(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

A03

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072761 A3

(51) International Patent Classification[7]: G01N 33/53, C12P 21/06, C12N 1/20, 15/00, 15/09, 15/63, 15/70, 15/74, C07K 1/00, 14/00, 17/00, C07H 21/04

(21) International Application Number: PCT/US02/06951

(22) International Filing Date: 7 March 2002 (07.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/274,526 8 March 2001 (08.03.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/274,526 (CIP)
Filed on 8 March 2001 (08.03.2001)

(71) Applicant *(for all designated States except US)*: **PHYLOS, INC.** [DE/DE]; 128 Spring Street, Lexington, MA 02421 (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HAMMOND, Philip, W.** [US/US]; 6 Hedgeway, Ayer, MA 01432 (US). **WRIGHT, Martin, C.** [GB/US]; 812 Memorial Drive #1105, Cambridge, MA 02139 (US). **ALPIN, Julia** [US/US]; 801 South Lincoln, Kirksville, MO 63501 (US). **FEKKES, Peter** [NL/US]; 53 Winslow Avenue, Somerville, MA 02144 (US).

(74) Agent: **EBLING, Karen, L.**; Clark & Elbing LLP, 101 Federal Street, Boston, MA 02110 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
6 March 2003

(48) Date of publication of this corrected version:
5 June 2003

(15) Information about Correction:
see PCT Gazette No. 23/2003 of 5 June 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/072761 A3

(54) Title: POLYPEPTIDES INTERACTIVE WITH BCL-X_L

(57) Abstract: Described herein are methods and reagents for identifying polypeptides that bind to a Bcl-X_L polypeptide, and methods for identifying compounds that modulate the interaction between a Bcl-X_L-binding polypeptide and a Bcl-X_L polypeptide.

POLYPEPTIDES INTERACTIVE WITH BCL-X_L

Background of the Invention

In general, the present invention relates to polypeptides that bind to Bcl-X_L, methods for identifying such polypeptides, and methods for identifying compounds that modulate the interaction between a Bcl-X_L-binding polypeptide and Bcl-X_L.

With the impending completion of the human genome sequence, interest is shifting to the emergent field of proteomics. One critical aspect of proteomics is the creation of a comprehensive map of protein-protein interactions. Such interactions are responsible for most signal transduction, making them attractive targets for drug therapy.

The primary methodology currently in use for interaction mapping is the yeast two-hybrid assay. Recently, genome-wide efforts to map protein-protein interactions have been reported for *S. cerevisiae* and, to a more limited extent, for *C. elegans* (Ito et al., Proc. Natl. Acad. Sci. U.S.A. 97:1143-1147, 2000; Uetz et al., Nature 403:623-627, 2000; and Walhout et al., Science 287:116-122, 2000). In the two-hybrid assay, the interaction of two proteins brings together their respective fusion partners, the DNA binding and activation domains of a transcription factor such as GAL4. This interaction thereby increases the transcription of a reporter gene that provides for the identification of interacting pairs.

While the yeast two-hybrid system has emerged as the leading technology in the field of protein-protein interactions, it is not without significant limitations. Firstly, the yeast two-hybrid system is limited by the *in vivo* nature of the assay. Binding interactions must take place under the conditions in the nucleus of the yeast cell, and many extracellular proteins are

unstable under these reducing conditions. In addition, proteins may prove toxic to the yeast through interactions with host cell proteins. Secondly, in order to generate a signal the two protein partners must be fused in an orientation that allows productive binding. Thirdly, because the two-hybrid system is a screening technique, there are practical limitations on the number of colonies that can be assayed.

Display technologies provide a powerful alternative and bypass many of the limitations of the two-hybrid system (Zozulya et al., Nat. Biotechnol. 17:1193-1198, 1999). In display methods, the interaction between a library member and a target polypeptide occurs *in vitro*, allowing optimal binding conditions to be used for different targets. Additionally, large libraries are screened iteratively, thus allowing even very low copy number proteins to be identified. However, in its most widely practiced form, phage display, this approach has similarly been hampered by the limitations of living systems. Specifically, libraries must be cloned, which decreases representation of the library members, can lead to the loss of sequences unstable in *E. coli*, and requires that proteins be properly processed to allow assembly of phage particles. In addition, the generation of libraries large enough to cover the entire proteome is difficult.

Summary of the Invention

The present invention features the application of mRNA display to the identification of protein-protein interactions involving the anti-apoptotic protein Bcl-X_L. The anti-apoptotic activity of Bcl-X_L is antagonized through binding to pro-apoptotic members of the Bcl-2 family, and protein members of the Bcl-2 family have been proposed as targets for drug therapy (Kinscherf et al., Expert. Opin. Investig. Drugs 9:747-764, 2000; Mattson and Culmsee, Cell Tissue Res. 301:173-187, 2000; and Chaudhary et al., Environ. Health

Perspect. 107 Suppl 1:49-57, 1999). Methods for identifying Bcl-X_L-binding polypeptides through mRNA display, as well as polypeptides identified as Bcl-X_L-binding polypeptides and the nucleic acid sequences encoding such polypeptides are described herein.

Accordingly, in a first aspect, the invention features a substantially pure human Bcl-X_L-binding polypeptide consisting of the sequence of any of SEQ ID NOS: 4-50, 63-71, and 224-228, or containing the sequence of any of SEQ ID NOS: 51-62, 229, and 230, as well as isolated nucleic acid molecules encoding those polypeptides (that is, SEQ ID NOS: 4-71 and 224-230), and vectors and cells containing those isolated nucleic acid molecules. In one embodiment, the nucleic acid molecule consists of the sequence of any of SEQ ID NOS: 156-202, 215-223, and 231-235. In another embodiment, the nucleic acid molecule contains the sequence of any of SEQ ID NOS: 203-214, 236, and 237. In another embodiment, the cell contains the vector into which an isolated nucleic acid molecule encoding a polypeptide of any of SEQ ID NOS: 4-71 and 224-230 is incorporated.

In a second aspect, the invention features a method of identifying a Bcl-X_L-binding polypeptide. The method involves providing a population of source labeled nucleic acid-protein fusion molecules; contacting the population of nucleic acid-protein fusion molecules with a Bcl-X_L polypeptide under conditions that allow interaction between the protein portion of a nucleic acid-protein fusion molecule of the population and the Bcl-X_L polypeptide; and detecting an interaction between the protein portion and the Bcl-X_L polypeptide, thereby identifying a Bcl-X_L-binding polypeptide. In a preferred embodiment, the population of source labeled nucleic acid-protein fusion molecules is derived from more than one source. In another preferred embodiment, the nucleic acid-protein fusion molecules are detectably-labeled. In yet another preferred embodiment, the Bcl-X_L polypeptide is immobilized

on a solid support, and the detection of an interaction between the protein portion of a nucleic acid-protein fusion molecule and a Bcl-X_L polypeptide is carried out by detecting the labeled nucleic acid-protein fusion molecule bound to the solid support. In this case, the support is preferably a bead or a chip.

In a third aspect, the invention features a method of identifying a compound that modulates binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide. The method entails contacting a Bcl-X_L polypeptide with (i) a Bcl-X_L-binding polypeptide consisting of the sequence of any of SEQ ID NOS: 4-50, 63-71, and 224-228, or containing the sequence of any one of SEQ ID NOS: 51-62, 229, and 230, and (ii) a candidate compound, under conditions that allow binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide. The level of binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide is then determined. An increase or decrease in the level of binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide, relative to the level of binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide in the absence of the candidate compound, indicates a compound that modulates the interaction between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide. The modulation may be an increase or a decrease in binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide.

In one embodiment of this aspect of the invention, the Bcl-X_L-binding polypeptide is part of a nucleic acid-protein fusion molecule. In a preferred embodiment, the Bcl-X_L-binding polypeptide is a free polypeptide that is not part of a fusion. In another preferred embodiment, the Bcl-X_L polypeptide is attached to a solid support. In yet another preferred embodiment, the Bcl-X_L-binding polypeptide is detectably-labeled, and the level of binding between the Bcl-X_L polypeptide and the Bcl-X_L-binding polypeptide is determined by measuring the amount of Bcl-X_L-binding protein

that binds to the solid support. Preferably, the solid support is a chip or a bead.

In a fourth aspect, the invention features a method of source-labeling a nucleic acid-protein fusion molecule. This method involves providing an RNA molecule; generating a first cDNA strand using the RNA molecule as a template; generating a second cDNA strand complementary to the first cDNA strand, the second cDNA strand further including a nucleic acid sequence that identifies the source of the RNA molecule; generating a source labeled RNA molecule from the double stranded cDNA molecule of the previous step; attaching a peptide acceptor to the source labeled RNA molecule generated in the previous step; and *in vitro* translating the RNA molecule to generate a source labeled nucleic acid-protein fusion molecule.

In a related aspect, the invention features a source-labeled nucleic acid-protein fusion molecule, where the nucleic acid portion of the fusion molecule contains a coding sequence for the protein and a label that identifies the source of the nucleic acid portion of the fusion molecule.

In another related aspect, the invention features a method of identifying the source of the nucleic acid portion of a nucleic acid-protein fusion molecule. The method includes providing a population of nucleic acid-protein fusion molecules, each molecule containing a source label that identifies the source of the nucleic acid portion of the fusion; and determining the identity of the source label, thereby identifying the source of the nucleic acid portion of a nucleic acid-protein fusion molecule. In preferred embodiments, the source label is cell type-specific, tissue-specific, or species-specific. In another preferred embodiment, the population of nucleic acid-protein fusion molecules contains subpopulations of nucleic acid-protein fusion molecules from a plurality of sources.

In any of the above aspects of the invention, the nucleic acid-protein fusion molecule is preferably an RNA-protein fusion molecule, for example, as described by Roberts and Szostak (Proc. Natl. Acad. Sci. U.S.A. 94:12297-302, 1997) and Szostak et al. (WO 98/31700; and U.S.S.N. 09/247,190), hereby incorporated by reference. Alternatively, the nucleic acid-protein fusion molecule is a DNA-protein fusion molecule, for example a cDNA-protein fusion molecule. Such molecules are described, for example, in U.S.S.N. 09/453,190 and WO 00/32823, hereby incorporated by reference.

By "nucleic acid-protein fusion molecule" is meant a nucleic acid molecule covalently bound to a protein. The nucleic acid molecule may be an RNA or DNA molecule, or may include RNA or DNA analogs at one or more positions in the sequence. The "protein" portion of the fusion is composed of two or more naturally occurring or modified amino acids joined by one or more peptide bonds. "Protein," "peptide,"and "polypeptide" are used interchangeably herein.

By "substantially pure polypeptide" or "substantially pure and isolated polypeptide" is meant a polypeptide (or a fragment thereof) that has been separated from components that naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the polypeptide is a Bcl-X_L-binding polypeptide that is at least 75%, more preferably, at least 90%, and most preferably, at least 99%, by weight, pure. A substantially pure Bcl-X_L-binding polypeptide may be obtained, for example, by extraction from a natural source (e.g., a cell), by expression of a recombinant nucleic acid encoding a Bcl-X_L-binding polypeptide, or by chemically synthesizing the polypeptide. Purity can be measured by any appropriate method, e.g., by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

A protein is substantially free of naturally associated components when it is separated from those contaminants that accompany it in its natural state. Thus, a protein that is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. Accordingly, substantially pure polypeptides not only include those derived from eukaryotic organisms but also those synthesized in *E. coli* or other prokaryotes.

By a "Bcl-X_L-binding polypeptide" is meant a polypeptide that interacts with a Bcl-X_L polypeptide or a fragment of a Bcl-X_L polypeptide. The interaction of a Bcl-X_L-binding polypeptide with a Bcl-X_L polypeptide can be detected using binding assays described herein, or any other assay known to one skilled in the art. In addition, a Bcl-X_L-binding polypeptide may be contained in the protein portion of a nucleic acid protein fusion molecule.

By a "Bcl-X_L polypeptide" is meant a polypeptide that is substantially identical to the polypeptide sequence of GenBank Accession Number: Z23115, or a fragment thereof. For example, a Bcl-X_L polypeptide may consist of amino acids 1 to 211 of GenBank Accession Number: Z23115.

By "substantially identical" is meant a nucleic acid molecule exhibiting at least 50%, preferably, 60%, more preferably, 70%, still more preferably, 80%, and most preferably, 90% identity to a reference nucleic acid sequence or polypeptide. For comparison of nucleic acid molecules, the length of sequences for comparison will generally be at least 30 nucleotides, preferably, at least 50 nucleotides, more preferably, at least 60 nucleotides, and most preferably, the full length nucleic acid molecule. For comparison of polypeptides, the length of sequences for comparison will generally be at least 10 amino acids, preferably, at least 15 nucleotides, more preferably, at least 20 amino acids, and most preferably, the full length polypeptide.

The "percent identity" of two nucleic acid or polypeptide sequences can be readily calculated by known methods, including but not limited to those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, and Devereux, eds., M. Stockton Press, New York, 1991; and Carillo and Lipman, SIAM J. Applied Math. 48: 1073, 1988.

Methods to determine identity are available in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux et al., Nucleic Acids Research 12(1): 387, 1984), BLASTP, BLASTN, and FASTA (Altschul et al., J. Mol. Biol. 215: 403 (1990). The well known Smith Waterman algorithm may also be used to determine identity. The BLAST program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, et al., NCBI NLM NIH Bethesda, MD 20894). Searches can be performed in URLs such as the following http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html; or http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi. These software programs match similar sequences be assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

By a "compound," "test compound," or "candidate compound" is meant a chemical molecule, be it naturally-occurring or artificially-derived, and includes, for example, peptides, proteins, synthetic organic molecules, naturally-occurring organic molecules, nucleic acid molecules, and components thereof.

By a "solid support" is meant any solid surface including, without limitation, any chip (for example, silica-based, glass, or gold chip), glass slide, membrane, bead, solid particle (for example, agarose, Sepharose, polystyrene or magnetic bead), column (or column material), test tube, or microtiter dish.

By a "microarray" or "array" is meant a fixed pattern of immobilized objects on a solid surface or membrane. As used herein, the array is made up of polypeptides immobilized on the solid surface or membrane. "Microarray" and "array" are used interchangeably. Preferably, the microarray has a density of between 10 and 1,000 objects/cm^2.

By "detectably-labeled" is meant any means for marking and identifying the presence of a molecule, e.g., an oligonucleotide probe or primer, a gene or fragment thereof, a cDNA molecule, or an antibody. Methods for detectably-labeling a molecule are well known in the art and include, without limitation, radioactive labeling (e.g., with an isotope such as ^{32}P or ^{35}S) and nonradioactive labeling (e.g., with a fluorescent label, such as fluorescein, or a chemiluminescent label).

By a "source label" is meant a nucleic acid sequence that is attached to a nucleic acid-protein fusion molecule. The source label identifies the origin of the nucleic acid portion of a nucleic acid-protein fusion molecule. For example, a source label can identify a specific cell type, tissue type, or species from which the nucleic acid portion of a nucleic acid-protein fusion molecule is derived. The source label also permits the selection of nucleic acid-protein fusion molecules from a particular source from a pool of nucleic acid-protein

fusion molecules from various sources. For example, a primer or a probe can be designed to detect the source label of nucleic acid-protein fusion molecules from a particular source, thereby allowing amplification or detection by hybridization of those particular fusion molecules. Such a primer or probe can also be designed for use as a handle for purification of a nucleic acid molecule or a nucleic acid-protein fusion molecule.

By "sequence cluster" is meant a group of sequences that form a continuous single sequence when their overlapping sequences are aligned. For example, a cluster sequence can be a set of sequences that each contain sequences in common with the other members of the sequence cluster. Sequence clusters can be formed using, for example, the computer program MacVector.

By "high stringency conditions" is meant conditions that allow hybridization comparable with the hybridization that occurs using a DNA probe of at least 500 nucleotides in length, in a buffer containing 0.5 M $NaHPO_4$, pH 7.2, 7% SDS, 1 mM EDTA, and 1% BSA (fraction V), at a temperature of 65°C, or a buffer containing 48% formamide, 4.8X SSC, 0.2 M Tris-Cl, pH 7.6, 1X Denhardt's solution, 10% dextran sulfate, and 0.1% SDS, at a temperature of 42°C (these are typical conditions for high stringency Northern or Southern hybridizations). High stringency hybridization is also relied upon for the success of numerous techniques routinely performed by molecular biologists, such as high stringency PCR, DNA sequencing, single strand conformational polymorphism analysis, and *in situ* hybridization. In contrast to Northern and Southern hybridizations, these techniques are usually performed with relatively short probes (e.g., usually 16 nucleotides or longer for PCR or sequencing, and 40 nucleotides or longer for *in situ* hybridization). The high stringency conditions used in these techniques are well known to those skilled in the art of molecular biology, and may be found, for example, in

Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY, 1998, hereby incorporated by reference.

By "transgene" is meant any piece of DNA that is inserted by artifice into a cell, and becomes part of the genome of the organism that develops from that cell. Such a transgene may include a gene that is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell that includes a DNA sequence that is inserted by artifice into a cell and becomes part of the genome of the organism that develops from that cell. As used herein, the transgenic organisms are generally transgenic mammals (e.g., mice, rats, and goats) and the DNA (transgene) is inserted by artifice into the nuclear genome.

By "knockout mutation" is meant an artificially-induced alteration in the nucleic acid sequence (created via recombinant DNA technology or deliberate exposure to a mutagen) that reduces the biological activity of the polypeptide normally encoded therefrom by at least 80% relative to the unmutated gene. The mutation may, without limitation, be an insertion, deletion, frameshift mutation, or a missense mutation. The knockout mutation can be in a cell *ex vivo* (e.g., a tissue culture cell or a primary cell) or *in vivo*.

A "knockout animal" is a mammal, preferably, a mouse, containing a knockout mutation as defined above.

By "transformation," "transfection," or "transduction" is meant any method for introducing foreign molecules into a cell, e.g., a bacterial, yeast, fungal, algal, plant, insect, or animal cell. Lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, retroviral delivery, electroporation, and biolistic transformation are just a few of the methods known to those skilled in the art which may be used. In addition, a foreign molecule can be introduced into a cell using a cell

penetrating peptide, for example, as described by Fawell et al. (Proc. Natl. Acad. Sci. U.S.A. 91:664-668, 1994) and Lindgren et al. (TIPS 21:99-103, 2000).

By "transformed cell," "transfected cell," or "transduced cell," is meant a cell (or a descendent of a cell) into which a nucleic acid molecule encoding a polypeptide of the invention has been introduced, by means of recombinant nucleic acid techniques.

By "promoter" is meant a minimal sequence sufficient to direct transcription. If desired, constructs of the invention may also include those promoter elements that are sufficient to render promoter-dependent gene expression controllable in a cell type-specific, tissue-specific, or temporal-specific manner, or inducible by external signals or agents; such elements may be located in the 5' or 3' or intron sequence regions of the native gene.

By "operably linked" is meant that a gene and one or more regulatory sequences are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequences.

By "sample" is meant a tissue biopsy, cells, blood, serum, urine, stool, or other specimen obtained from a patient or test subject. The sample is analyzed to detect a mutation in a gene encoding a Bcl-X_L-binding polypeptide, or expression levels of a gene encoding a Bcl-X_L-binding polypeptide, by methods that are known in the art. For example, methods such as sequencing, single-strand conformational polymorphism (SSCP) analysis, or restriction fragment length polymorphism (RFLP) analysis of PCR products derived from a patient sample may be used to detect a mutation in a gene encoding a Bcl-X_L-binding polypeptide; ELISA may be used to measure levels of a Bcl-X_L-binding polypeptide; and PCR may be used to measure the level of nucleic acids encoding a Bcl-X_L-binding polypeptide.

By "apoptosis" is meant cell death characterized by any of the following properties: nuclear condensation, DNA fragmentation, membrane blebbing, or cell shrinkage.

By "modulating" is meant either increasing ("upward modulating") or decreasing ("downward modulating") the number of cells that undergo apoptosis in a given cell population. Preferably, the cell population is selected from a group including cancer cells (e.g., ovarian cancer cells, breast cancer cells, pancreatic cancer cells), leukemic cells, lymphoma cells, T cells, neuronal cells, fibroblasts, or any other cell line known to proliferate in a laboratory setting. It will be appreciated that the degree of apoptosis modulation provided by an apoptosis modulating compound in a given assay will vary, but that one skilled in the art can determine the statistically significant change in the level of apoptosis that identifies a compound that increases or decreases apoptosis. Preferably, for downward modulating, apoptosis is decreased by least 20%, more preferably, by at least, 40%, 50%, or 75%, and, most preferably, by at least 90%, relative to a control sample which was not administered an apoptosis downward modulating test compound. Also as used herein, preferably, for upward modulating, apoptosis is increased by at least 1.5-fold to 2-fold, more preferably, by at least 3-fold, and most preferably, by at least 5-fold, relative to a control sample which was not administered an apoptosis upward modulating test compound.

By an "apoptotic disease" is meant a condition in which the apoptotic response is abnormal. This may pertain to a cell or a population of cells that does not undergo cell death under appropriate conditions. For example, normally a cell will die upon exposure to apoptotic-triggering agents, such as chemotherapeutic agents, or ionizing radiation. When, however, a subject has an apoptotic disease, for example, cancer, the cell or a population of cells may not undergo cell death in response to contact with apoptotic-

triggering agents. In addition, a subject may have an apoptotic disease when the occurrence of cell death is too low, for example, when the number of proliferating cells exceeds the number of cells undergoing cell death, as occurs in cancer when such cells do not properly differentiate.

An apoptotic disease may also be a condition characterized by the occurrence of inappropriately high levels of apoptosis. For example, certain neurodegenerative diseases, including but not limited to Alzheimer's disease, Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis, multiple sclerosis, restenosis, stroke, and ischemic brain injury are apoptotic diseases in which neuronal cells undergo undesired cell death.

By "proliferative disease" is meant a disease that is caused by or results in inappropriately high levels of cell division, inappropriately low levels of apoptosis, or both. For example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, and lung cancer are all examples of proliferative disease.

By a "substantially pure nucleic acid," "isolated nucleic acid," or "substantially pure and isolated nucleic acid" is meant nucleic acid (for example, DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid of the invention is derived, flank the nucleic acid. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

By "antisense," as used herein in reference to nucleic acids, is meant a nucleic acid sequence, regardless of length, that is complementary to the coding strand of a nucleic acid molecule encoding a Bcl-X_L polypeptide or a Bcl-X_L-binding polypeptide. Preferably, the antisense nucleic acid molecule is capable of modulating apoptosis when present in a cell. Modulation of at least 10%, relative to a control, is recognized; preferably, the modulation is at least 25%, 50%, or more preferably, 75%, and most preferably, 90% or more.

By a "purified antibody" is meant an antibody that is at least 60%, by weight, free from proteins and naturally occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably, 90%, and, most preferably, at least 99%, by weight, antibody, e.g., a Bcl-X_L-binding polypeptide-specific antibody. A purified antibody may be obtained, for example, by affinity chromatography using recombinantly-produced protein or conserved motif peptides and standard techniques.

By "specifically binds" is meant a compound that recognizes and binds a protein or polypeptide, for example, a Bcl-X_L polypeptide or a Bcl-X_L-binding polypeptide, and that when detectably labeled can be competed away for binding to that protein or polypeptide by an excess of compound that is not detectably labeled. A compound that non-specifically binds is not competed away by the above excess detectably labeled compound.

The present invention has several utilities. Since the Bcl-2 family of proteins, and Bcl-X_L itself, has been implicated in apoptosis, these Bcl-2 family polypeptides can be used in screens for therapeutics that modulate diseases or developmental abnormalities involving overactivity or underactivity of apoptotic pathways. In particular, Bcl-X_L is known to protect cancer cells (e.g., pancreatic carcinoma cells) from stimulation of apoptosis, and this effect is reversible by adding an agent, Bax (Hinz et al., Oncogene 19:5477-5486, 2000), that binds to Bcl-X_L at the same site as many of the polypeptides of the

present invention. Therefore, the polypeptides that bind to Bcl-X_L, described herein, may be used as targets in therapeutics screening assays. The identified polypeptides are particularly useful in such screens because they represent the functional portions of human proteins that bind Bcl-X_L. These polypeptides may also be used to detect Bcl-X_L polypeptides in a sample. In addition, the methods of the present invention are useful as high-throughput screening methods for potential therapeutics involved in the overactivity or underactivity of apoptotic pathways.

The general approach of the present invention also provides a number of advantages. For example, direct mRNA display allows the mapping of protein-protein interactions, which is useful for drug screening. In mRNA display (Roberts and Szostak, supra), a DNA template is used to transcribe an engineered-mRNA molecule possessing suitable flanking sequences (e.g., a promoter; a functional 5' UTR to allow ribosome binding; a start codon; an open reading frame; a sequence for polypeptide purification; and a conserved sequence used for ligation to a complementary linker containing puromycin). To the 3' end of the mRNA, a linker strand with a puromycin moiety (Pu) is then added, preferably by photo-crosslinking. When this RNA is translated *in vitro*, the puromycin becomes incorporated at the C-terminus of the nascent peptide. The resulting mRNA display construct is then purified after ribosome dissociation. A cDNA strand is then synthesized to protect the RNA and to provide a template for future PCR amplification. A library of such constructs can be incubated with immobilized target, and molecules that bind are enriched by washing away unbound material. Bound cDNAs are recovered, for example, by KOH elution, and subsequent PCR is performed to regenerate a library enriched for target-binding peptides. Figure 1 shows the steps involved in mRNA display.

As mRNA display is a completely *in vitro* technique, many of the problems inherent in cloning and expression are eliminated. The elimination of cloning bottlenecks in library preparation allows the generation of very large libraries, routinely in the range of 10^{13} members. In addition, the formation of mRNA display constructs is readily achieved in a mammalian expression system, thereby providing suitable chaperones for the folding of human proteins and the potential for appropriate post-translational modifications.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

FIGURE 1 is a schematic representation of iterative selection using mRNA display.

FIGURE 2A shows the sequences of positive control polypeptides used in Bcl-X_L polypeptide binding assays (SEQ ID NOS: 238-240).

FIGURE 2B is a graph showing the binding of control polypeptides to a Bcl-X_L polypeptide.

FIGURE 3A shows polypeptides identified as Bcl-X_L binding polypeptides using methods described herein (SEQ ID NOS: 1-71), as well as information on the binding affinity and specificity of the polypeptides. In addition, the number of clones of each sequence cluster obtained from each library is presented.

FIGURE 3B shows the polypeptide sequences of Bcl-X_L-binding polypeptides (SEQ ID NOS: 1-71), and indicates corresponding nucleic acid sequences.

FIGURE 3C shows the nucleotide sequences of selected Bcl-X_L-binding polypeptides (SEQ ID NOS: 72-142). The nucleotide sequences encoding the selected Bcl-X_L-binding polypeptides are underlined (SEQ ID

NOS: 153-223).

FIGURE 4 is a schematic representation of the alignment of selected Bcl-X_L-binding polypeptides within their parental proteins. Each unique fragment was analyzed to determine the location of the amino- and carboxyl-termini within the parental protein sequence, and these amino acids are indicated by residue and number. The number of isolated clones corresponding to each unique fragment was determined and is indicated next to the fragment ID. These fragments are mapped against the parental sequences of Bim, Bax, HSPC300, and TPR (SEQ ID NOS: 241-244). The BH3 domain core sequence is underlined for the BimL and Bax proteins. Splice variants are indicated by a * in the ID and the use of (-) in place of (=) in the fragment map.

FIGURE 5 is a graph of the relative binding affinity of a selected Bak Bcl-X_L-binding polypeptide to immobilized Bcl-X_L polypeptide versus concentration of immobilized Bcl-X_L polypeptide.

FIGURE 6 is a graph of the effect of the binding of a Bcl-X_L-binding polypeptide in the presence of a competitor BH3 domain from the Bcl-2 family member Bak.

FIGURE 7 shows the polypeptide sequences of representative clones (SEQ ID NOS: 1, 5, 245, 60, 61, 6, 46, 2, 33, 4, 7-10, 3, 11, 48, 12, 53, and 54) from sequence clusters that were bound to a Bcl-X_L polypeptide in the presence of a competitor BH3 domain from the Bcl-2 family member Bak. Competitive binding was determined relative to a control containing no competitor. The selected polypeptide sequences are shown aligned by sequence homology, where possible, to the known BH3 domains of Bim, Bak, and Bax.

FIGURES 8A, 8B, and 8C are tables of amino acid sequences that bind Bcl-X_L protein (FIGS. 8A and 8B; SEQ ID NOS: 224-230) and their nucleic acid coding sequences (FIG 8C; SEQ ID NOS: 231-237).

FIGURE 9 is a graph showing free peptide binding to GST-BCL-X_L, as compared to background binding to GST. Bax was used as a positive control for BCL-X_L binding.

Described herein are methods for identifying polypeptides that interact with a Bcl-X_L polypeptide; methods for identifying compounds that increase or decrease the binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide; methods for source labeling a nucleic acid-protein fusion molecule; and methods for identifying the source of the nucleic acid portion of a nucleic acid-protein fusion molecule. Techniques for carrying out each method of the invention are now described in detail, using particular examples. The examples are provided for the purpose of illustrating the invention, and should not be construed as limiting. Also described herein are novel Bcl-X_L-binding polypeptides and nucleic acid molecules obtained through the methods of the present invention.

Materials and methods for identifying Bcl-X_L-binding polypeptides

The experiments described herein were carried out using the materials and methods described below.

Choice of UTR sequence tags

Unique UTR sequences that are compatible with translation in rabbit reticulocyte lysate were identified by selection from a library of c-*myc* mRNAs with a partially randomized 5' UTR. The c-*myc* construct described by Roberts and Szostak (supra) was amplified by PCR using the 5' primer TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT HHH HHH HHA CAA TGG CTG AAG AAC AGA AAC TG (where H is an equimolar mixture of A, C, and T) (SEQ ID NO: 143). This amplification inserted 8 random bases into the 5' UTR upstream of the ATG start codon, to

give a library of 3^8 (6561) different mRNA molecules after *in vitro* transcription with T7 RNA polymerase. Fusion formation, reverse transcription, and immunoprecipitation with an anti-c-*myc* antibody were carried out as described by Roberts and Szostak (supra) to separate mRNAs that had undergone translation from those that had not. The successfully translated and fused sequences were amplified by PCR using the 5' primer TAATACGACTCA CTATAGGGACAATTACTATTTACAATT (SEQ ID NO: 144), in which the T7 promoter is underlined, to preserve the information in the randomized region. Sequences obtained from individual clones were subsequently used in the construction of tagged libraries.

Library preparation

The design of the above-described sequence tags can be adapted to source label nucleic acid sequences from various sources. Instead of each sequence tag being a unique sequence (as described above), one sequence tag (source label) is used to label a set of nucleic acid sequences derived from the same cell, tissue, or species. The source labeled sequences can then be pooled with different source labeled sequences and used for mRNA display as described herein, and the origin of each sequence in the pool can be determined.

Individual RNA sequences are translated *in vitro*, and RNA-protein fusions are formed, for example, according to the methods of Roberts and Szostak (supra) and Szostak et al. (WO 98/31700; U.S.S.N. 09/247,190), hereby incorporated by reference. Specifically, each mRNA display library was prepared according to the following methods. Poly-A+ mRNA (Clontech) was primed using the oligonucleotides GGAACTTGCTTCGTCTTTGCAATCN$_9$ (SEQ ID NO: 145) or GGATGATGCTTCGTCTTTGTAATCN$_9$ (SEQ ID NO: 146) and a cDNA

molecule was synthesized using SuperScript II Reverse Transcriptase (Promega). Two primers were used initially, to allow the investigation of different ligation sequences; these sequences were subsequently altered and made uniform by the use of a single PCR primer under conditions that would allow it to anneal to either template. The RNA/cDNA hybrid molecule was then treated with RNase H in order to partially degrade the RNA member of the hybrid molecule. Unextended primers were then removed by purification over an S-300 (Pharmacia) size exclusion column.

Second strand cDNA synthesis was carried out by the Klenow fragment of *E. coli* DNA polymerase, using primers having the sequence GGACAATTACTATTTACAATT[H_8]ACAATGN$_9$ (SEQ ID NO: 147) that included a 5' UTR with a sequence tag H_8 (source label), derived as described above, and a start codon (underlined). In the production of libraries from human kidney, liver, bone marrow, and brain mRNAs, the source labels CTCCTAAC (SEQ ID NO: 250), CTTTCTCT (SEQ ID NO: 251), CTTACTTC (SEQ ID NO: 252), and ATTTCAAT (SEQ ID NO: 253) were used, respectively. Unextended primers were again removed by S-300 size exclusion chromatography, and the cDNA product was then PCR amplified using a forward primer encoding the T7 promoter (underlined) and 5' UTR, TAATACGACTCACTATAGGGACAATTACTATTTACAATT (SEQ ID NO: 148), and reverse primers corresponding to the fixed regions of the first strand primers above. After PCR product purification using spin columns (Qiagen), short fragments were removed by S-300 size exclusion chromatography.

mRNA display construct formation

The above described PCR products were reamplified using the forward primer described above

(TAATACGACTCACTATAGGGACAATTACTATTTACAATT) (SEQ ID NO: 148) and a single reverse primer, TTTTAAA*TAGCGCATGC*CTTATCGTCATCGTCTTTGTAATC (SEQ ID NO: 149), encoding the FLAG-M2 epitope (underlined) and a region complementary to the photoligation linker (italics). The single reverse primer was used to amplify libraries containing each of the initial first strand primer sequences in order to produce a single uniform end. These amplicons were then used as templates for transcription using T7 RNA polymerase (Ambion MegaScript). The resulting RNA molecules were purified by phenol/chloroform/isoamyl alcohol extraction and NAP column (Pharmacia) purification. The puromycin-containing linker 5'-Pso-TAGCGGATGCA$_{18}$XXCCPu (where X is PEG spacer 9; Pso is psoralen; and Pu is puromycin) was photo-ligated to the 3' end of the RNA essentially as described by Kurz et al. (Nucleic Acids Res. 28:E83, 2000). Ligated RNA molecules were then translated for 30 min at 30°C in a 300 µl reaction containing 200 µl of rabbit reticulocyte lysate (Ambion), 120 pmole of ligated RNA, 10 µl of an amino acid mix lacking methionine (Ambion), and 15 µl of ^{35}S-met (Amersham). Subsequently, 100 µl of 2M KCl and 25 µl of 1M $MgCl_2$ was added to facilitate formation of the mRNA display complex. The mRNA display constructs were then purified by binding to 100 µl of 50% oligo-dT cellulose slurry in a total volume of 10 ml (100 mM Tris-HCl (pH 8), 10 mM EDTA, 1M NaCl, and 0.25% Triton X-100) at 4°C for 1 hr. The binding reaction was then transferred to a column (BIORAD), washed 3 times with 1 ml of binding buffer containing no EDTA, then eluted with 100 µl aliquots of 2 mM Tris-HCl (pH 8), 0.05% Triton X-100, and 5 mg/ml BSA.

A cDNA strand was synthesized using SuperScript II RT (Promega) and the reverse sense PCR primer in the manufacturer supplied buffer. The reverse transcription reaction was then diluted to 1 ml in TBK buffer (50 mM

Tris-HCl (pH 7.5), 150 mM KCl, and 0.05% Triton X-100) and incubated with 200 µl of anti-FLAG Ab immobilized on agarose beads (Sigma) for 1 hr at 4°C. The binding reaction was transferred to a column and the beads were washed 3 times with 1 ml of TBK buffer. mRNA-display constructs were then eluted with 100 µl aliquots of TBK buffer containing 100 µM FLAG-M2 peptide, 5 mg/ml BSA, and 0.1 mg/ml salmon sperm DNA. The yield of mRNA-protein fusion product was determined by scintillation counting the purified product and comparing it to an estimated specific activity of methionine based on an approximate concentration of 5 µM in the lysate. For the libraries containing a heterogeneous population of proteins, the prevalence of methionine was approximated as one initiator methionine per molecule plus one for each 60 amino acids.

Target protein preparation

A portion of the human Bcl-X_L gene was PCR amplified from a GeneStorm® Expression-Ready Bcl-X_L clone (Invitrogen, Carlsbad, CA) using the primers AGTATCGAATTCATGTCTCAGAGCAACCGG (SEQ ID NO: 150) and TACAGTCTCGAGCTAGTTGAAGCGTTCCTGGCCCT (SEQ ID NO: 151). The 644 nucleotide Bcl-X_L DNA fragment obtained from the above PCR reaction was then cloned into the expression vector 4T-1 (Pharmacia). Competent *E. coli* (BL21(DE3) pLysS) were transformed with the expression vector and grown on LB/carbenicillin plates overnight at 37°C. A single transformed colony was then selected and grown overnight in 5 ml of LB/carbenicillin. Two ml of this starter culture was used to inoculate a fresh 100 ml culture, which was grown at 37°C until an OD_{600} of 0.6 was reached.

Expression of the Bcl-X_L polypeptide was induced in the bacterial culture by the addition of IPTG to a final concentration of 0.4 mM, and the culture was shaken at 25°C overnight. The bacterial cells were then harvested

for their Bcl-X_L polypeptide by centrifugation at 12,000g for 30 minutes. The cell pellet was resuspended in 1/10 volume 100 mM Tris/HCl (pH 8.0)/ 100 mM NaCl/ 0.1% Triton X-100/ 1.0% glycerol, and the cells were lysed by dounce homogenization and three freeze/thaw cycles. The bacterial cell lysate was clarified by centrifugation at 16,000g for 30 minutes, and 5 ml of the clarified lysate was applied to a 2 ml RediPack glutathione column (Pharmacia). The column was washed with 20 ml of lysis buffer and eluted, in a stepwise manner, with lysis buffer to which reduced glutathione had been added, to final concentrations of 1, 5, 10, 15, and 20 mM. Fractions of the eluate were analyzed on 4-12% NuPAGE gels (Novex) and positive fractions, based on polypeptide size, were pooled. The protein was dialyzed against 100 mM Tris/HCl (pH 8.0)/ 100 mM NaCl/ 0.05% Triton X-100/ 1.0 % glycerol and the protein concentration was determined by BCA assay (Pierce).

Assay to Detect Binding of a Polypeptide to a Bcl-X_L-GST Fusion Protein

Detection of a polypeptide binding to a Bcl-X_L-glutathione S-transferase (GST) fusion protein was carried out as follows. Twenty microliters of glutathione Sepharose 4B slurry (AP Biotech) was aliquoted to a microcentrifuge tube and washed with PBS. The Bcl-X_L-GST fusion protein (60 μg), prepared as describe above, was added and allowed to bind to the Sepharose beads for 1 hr at 4°C. The beads were then re-washed in selection buffer (50 mM Tris-HCL pH 7.5, 150 mM KCl, 0.05% Triton X-100, 0.5 mg/ml BSA, and 0.1 mg/ml salmon sperm DNA). The Bcl-X_L-GST beads were resuspended in 100 μl of selection buffer (approximately 11.5 μM Bcl-X_L) and ^{35}S-labeled mRNA display construct or free peptide was added (approximately 10-60 nM) and incubated on a rotator for 1 hr at 4°C. The reaction was then transferred to a microcentrifuge column (BioRad) and unbound mRNA display constructs or free peptides were removed by a 10 sec

spin at 1,000 rpm. The Sepharose beads were then washed three times with 500 µl of selection buffer. The extent of binding between the Bcl-X_L-GST fusion protein and the mRNA display constructs or free peptides was determined by scintillation counting each fraction, including the recovered beads.

Selection

A human Bcl-X_L-GST fusion protein was immobilized on Sepharose beads as described above for the binding assay and incubated with the mRNA display library. For the first round of selection, the input was approximately 0.06 pmol of each of the four source labeled libraries from human tissues (kidney, liver, bone marrow, and brain), which were mixed prior to selection. For subsequent rounds of selection, the input of each of the four source labeled libraries ranged from 0.25 to 0.92 pmol in total. After washing the beads of any unbound nucleic acid-protein fusion library members, the cDNA strand of the bound fusions were recovered in three elutions with 100 µl of 0.1N KOH. Eluates were subsequently neutralized by the addition of 2 µl of 1M Tris-HCL pH 7 and 8 µl of 1N HCL. A small scale PCR optimization was performed with the eluate to determine the number of cycles required to produce a strong signal without overamplification (typically 18-28 cycles). The library was then regenerated by PCR using the remainder of the eluate.

Cloning and sequencing of library members that bind to the Bcl-X_L-GST fusion protein

PCR products of the selected library members that bound to the Bcl-X_L-GST fusion protein were cloned into the TOPO-TA vector (Invitrogen) and, after isolation of individual colonies, the plasmids were purified (Qiagen) and sequenced using standard sequencing techniques (Ausubel et al., supra).

In vitro synthesis of polypeptides that bind to the Bcl-X_L-GST fusion protein

To synthesize polypeptides that interacted with the Bcl-X_L-GST fusion protein, RNAs were prepared from the PCR products of the selected library members that bound to the Bcl-X_L-GST fusion protein and purified as described above. After translation in rabbit reticulocyte lysate (Ambion), the peptides were purified directly from the lysate by immunoprecipitation and peptide elution based on a C-terminal FLAG-M2 epitope contained in the peptide (Sigma).

Detection of known Bcl-X_L-binding polypeptides

Members of the Bcl-2 family of apoptotic proteins function via homo- and heterodimerization, occurring primarily through the binding of a single α-helix designated the BH3 domain (Bcl-2 Homology domain 3) in a corresponding pocket produced by three α-helices in the interacting partner (Diaz et al., J. Biol. Chem. 272:11350-11355, 1997; and Sattler et al., Science 275:983-986, 1997). The target protein used herein was the human Bcl-X_L protein produced as a GST fusion and immobilized on glutathione Sepharose beads. The BH3 domains of three different Bcl-2 family proteins (Bcl-2, Bax, and Bak) were prepared as mRNA display constructs, as described herein, along with control peptides derived from unrelated proteins Stat-1 and Raf-1. The BH3 domains of Bcl-2, Bax, and Bak are shown with the consensus regions aligned and highlighted in Figure 2A. Individual mRNA display constructs were incubated with either the target Bcl-X_L-GST fusion protein bound to glutathione beads or with the beads alone. Unbound materials were collected, and the beads were washed. The amount of peptide bound to the beads was determined by scintillation counting and graphed as the percent of input counts bound (Figure 2B).

Binding of Bcl-2, Bax, and Bak to Bcl-X_L-GST fusion protein was specific to the BH3 helices, with Bak binding most efficiently (40%) followed by Bax (6%); no binding was observed for the BH3 helix from Bcl-2 or either control. The ordering of Bax and Bak is in good agreement with published IC_{50} values which indicate that Bcl-X_L has an affinity for the Bak BH3 domain that is approximately five-fold higher than that for Bax (Diaz et al., supra). The lack of binding observed for the BH3 domain of Bcl-2 could be due to the BH3 domain peptide failing to form a helix (Zhang et al., Biochem. Biophys. Res. Commun. 208:950-956, 1995; and Xie et al., Biochemistry 37:6410-6418, 1998), or the affinity may be below that required to generate a signal in this assay.

Identification of novel Bcl-X_L-binding polypeptides

Having established the binding of Bcl-X_L control peptides, as described above, a selection to identify binders from within the complex mixture of an mRNA display library was initiated (Figure 1). Four libraries, individually prepared from the tissue-specific mRNAs of kidney, liver, bone marrow, and brain were pooled prior to initiating selection. Each library contained a unique 8 nucleotide (nt) tag (source tag) within the 5' UTR to allow specific amplification of an individual library. The ability to mix libraries not only increased the size and diversity of the starting pool, but the identification of tissue of origin for each selected protein provided information similar to that normally obtained from mRNA expression analysis.

As a target for the selection, a GST fusion protein of Bcl-X_L was immobilized on glutathione Sepharose beads. The selection was initiated with a combined library of approximately $1.5x10^{11}$ molecules. After incubation of the library with the target, unbound members of the library were washed away and the bound material was eluted. An enriched library was then regenerated

by PCR, transcription, ligation, translation, fusion, reverse-transcription, and purification. This enriched library was then used for the subsequent round of selection.

After four rounds of selection, the enriched pool from the combined libraries bound the Bcl-X_L target at about 40%, an extent similar to the Bak control construct (see Figure 2B). In order to determine if the selected Bcl-X_L polypeptides originated from one or multiple libraries, each library was prepared individually after specific amplification using library specific primers. The library constructed from brain mRNA was omitted due to cross-reaction of the PCR amplification primer. A test of binding revealed that each tissue-specific library bound to the target to an extent similar to the mixed pool. The bound material from each of the individual libraries was then recovered by elution, PCR amplified, and analyzed by cloning and sequencing.

Additional rounds of selection may change the population distribution significantly. A rare sequence from the starting pool that binds tightly might be enriched only to the point of appearing once among the clones while a poorer binding sequence that was abundant in the starting pool might still be found at high copy number. Also, sequencing more clones may lead to the identification of other proteins still present at low copy number.

Sequence analysis of Bcl-X_L-binding polypeptides

A total of 378 sequences were obtained from the above-described binding assay. Of the sequences, 181 were from the kidney library, 85 were from the liver library, and 112 were from the bone marrow library. Initial analysis of the sequences revealed a total of 71 distinct sequence clusters. Six of the clusters (8%) originated from all three libraries, 14 clusters (20%) originated from two of the three libraries, and the remaining 51 clusters (72%) originated from only one library. Many of the clusters contained a number of

identical clones as well as a variety of clones with distinct 5' or 3' ends. This variety reflects the random priming used to prepare the library and allowed minimal functional regions of the Bcl-X_L-binding polypeptides to be delineated based on the overlapping regions of individual family members (Figure 4). The sequences were then subjected to both nBLAST and pBLAST searches to identify the proteins represented by each cluster. Thirty-six of the clones were from known polypeptides (SEQ ID NOS: 1-28, 63-69, and 71), twenty-three of the clones were from hypothetical or unknown polypeptides whose nucleic acid sequences were found in the database (SEQ ID NOS: 29-50, and 70), and twelve clones were unique polypeptide sequences (SEQ ID NOS: 51-62). These Bcl-X_L-binding polypeptide sequences are shown in Figure 3B, and their corresponding nucleic acid sequences are shown underlined in Figure 3C.

Twenty of the most frequently found Bcl-X_L-binding polypeptides are provided in Table 1. The number of clones in each cluster was further broken down by the number containing the source label of each individual library (NF indicates none found among the clones sequenced). The identification number of the specific clone from each cluster chosen for further characterization is also indicated. The numbers present in Table 1 reflect the diversity of polypeptides that interact with other polypeptides attainable from large libraries generated by the *in vitro* methods of the invention.

Table 1. Frequently found Bcl-X_L binding polypeptides

Protein	Kidney (181)	Liver (85)	Marrow (112)	Total (378)	Clone ID
Bim	43	11	36	90	T44
HSPC300	9	15	11	35	C68
TPR, nuclear pore complex-associated protein	23	NF	NF	23	C55
Bax	19	NF	3	22	C49
Novel Protein A	1	11	6	18	V18
cDNA FLJ23277, Clone HEP03322	12	2	2	16	X42
Hypothetical protein DKFZp586HO623	NF	1	15	16	V47
Syntaxin 4A	8	4	NF	12	U58
Tumor protein HDCMB21P	1	5	5	11	V50
Proline/Glutamine rich splicing factor	7	1	1	9	--
Novel Protein B	3	5	NF	8	V68
Talin	4	NF	1	5	X56
Thyroid hormone receptor-associated protein	5	NF	NF	5	U25
Sterol regulatory element binding txn factor	NF	NF	5	5	W17
Bcl-2 related proline-rich protein BPR	NF	2	3	5	Y75
cDNA FLJ22171, clone HRC00654	NF	NF	5	5	T42
Toll-like receptor 3	4	NF	NF	4	U15
Calpain	1	3	NF	4	V53
Bak	2	1	NF	3	C32
Novel protein D	NF	1	2	3	T25

The most abundant Bcl-X_L-binding polypeptide (~25% of the total) was that of Bim, which was originally identified as a partner of Bcl-2 in a protein interaction screen and subsequently shown to bind to Bcl-X_L (O'Connor et al., EMBO J. 17:384-395, 1998). Two other proteins out of the top twenty, Bak and Bax, contain BH3 domains known to interact with Bcl-X_L (Diaz et al., supra). A fourth member of the Bcl-2 family, BPR, was also found in this screen. This newly reported member of the Bcl-2 family was not present in the database during the initial search. That a protein that was initially categorized as unknown is indeed a member of the Bcl-2 family reinforces the hypothesis that other novel polypeptides identified in the screen may also be members of the Bcl-2 family. While initial reports indicate that BPR contains a BH2 domain (Scorilas et al., unpublished, 2000), the present invention indicates that it also contains a BH3 domain.

Further analysis of the known Bcl-X_L-binding polypeptides was done to determine whether each selected Bcl-X_L-binding polypeptide sequence was from the coding region or UTR and if the reading frame matched that of the native protein. This analysis was used as a filter to eliminate false positives; polypeptides that failed at this step were not further characterized. Twenty-seven out of the thirty-six clusters from known polypeptides were in frame and within their native ORFs. Three out of thirty-six, proline/glutamine rich splicing factor (SEQ ID NO: 63), UDP glucoronosyl transferase 2B4 precursor (SEQ ID NO: 71), and cDNA FLJ20617 (SEQ ID NO: 70) were from the incorrect reading frame. Two clusters, transforming growth factor and arsenate resistance protein (SEQ ID NOS: 64 and 66, respectively), had inserts in the reversed orientation relative to the parent mRNA and probably arose due either to incomplete removal of the first strand primer after cDNA synthesis or re-priming on the cDNA strand after first strand synthesis. An additional four clusters were derived from reportedly noncoding regions of the parent mRNA, that is, the 3' UTR (L-plastin, K-ras oncogene, lysosomal pepstatin insensitive protease, and MYBPC3; SEQ ID NOS: 65, 67, 68, and 69, respectively).

Figure 4 shows an alignment of selected Bcl-X_L-binding polypeptides with their parental proteins, identified as described above. Each unique fragment was analyzed to determine the location of the amino and carboxyl termini within the parental protein sequence and these amino acids are indicated by residue and number. The number of isolated clones corresponding to each unique fragment was determined and is indicated next to the fragment ID. These fragments are mapped against the parental sequences of BimL, Bax, HSPC300, and TPR.

Affinity and specificity of the Bcl-X_L binding polypeptides

The initial sequencing data showed the relative frequency of each clone in the selected pool. Additional ranking of individual clones may provide valuable insight into the biological relevance of each interaction. For example, a binding affinity consistent with the cellular concentrations of the interacting proteins has been proposed as a litmus test for biological significance (Mayer, Mol. Biotechnol. 13:201-213, 1999). The great flexibility and precise control over assay conditions, such as target concentration and the presence of additives, is one of the advantages of the *in vitro* selection methods of the present invention. By ranking the selected polypeptides based on readily assayable characteristics, it is possible to quickly identify a subset of polypeptides for assays that address the *in vivo* activity of the identified polypeptides.

To determine the affinity of the selected Bcl-X_L-binding polypeptides, each cluster of selected sequences was aligned and the shortest sequence was generally chosen as representing the minimal binding domain for that particular cluster. It should be noted that this shortest fragment may represent only a partial binding sequence and longer fragments may bind with higher affinity. The chosen clones were prepared as free peptides and used in the binding assay described below.

Purified radioactively labeled protein from the individual clones was incubated with immobilized Bcl-X_L-GST for one hour and, after washing, the amount bound was determined by scintillation counting. The binding at each concentration was normalized to that at the highest concentration and plotted versus concentration. Figure 5 is a representational plot of the results of this binding assay. A selected Bak fragment (MGQVGRQLAIIGDDINRDYKDDDDKASA; SEQ ID NO: 152), containing a FLAG-M2 epitope, was synthetically produced as a free protein and used in a

binding assay in which the concentration of immobilized Bcl-X_L-GST was varied from 11 nM to 28 μM. The amount of peptide bound to Bcl-X_L was determined by scintillation counting and normalized to that bound at the highest concentration. Normalized binding was then plotted versus Bcl-X_L concentration and fit to a binding curve using nonlinear regression. In this assay, all of the clones except one showed binding that was clearly dependent on target concentration. However, only binding curves that gave a high correlation coefficient ® value) were used to determine an affinity.

Binding affinities of the free Bcl-X_L-binding polypeptides (i.e., Bcl-X_L-binding polypeptides that are not part of fusions) ranged from approximately 2 nM to 10 μM, demonstrating the great range of affinities accessible by *in vitro* selection. The twenty clones with the highest affinity are presented in Table 2. The indicated clone from each sequence cluster was produced *in vitro* and the relative K_d was determined for binding to Bcl-X_L. The total number of clones in that sequence cluster is indicated for comparison of affinity to abundance.

Table 2. High Affinity Bcl-X_L-binding polypeptides

clone ID	Protein	Accession Number	K_d (μM)	Total clones
T44	Bim	NP_006529	0.002	90
T95	Neutrophil cytosolic factor 2	NP_000424	0.00416	2
V47	Hypothetical protein DKFZp586ho623	NM_017540	0.0129	16
C21	Novel protein I	--	0.07	3
V18	Novel protein A	--	0.086	3
X56	Talin (splice variant)	NP_006280	0.093	6
V72	unknown protein from clone 425C14 on chrom. 6q22	Z99129	0.28	1
C32	Bak	NP_001179	0.402	3
Y37	unknown protein from cDNA: FLJ21691, clone COL09555	AK025344	0.41	1
Y75	Bcl-2 related protein BPR	AF289220	0.42	5
V06	Golgi SNAP receptor complex member 1	NP_004862	0.467	1
C68	HSCP300	AF161418	0.58	35
U58	Syntaxin	NP_004595	0.64	12
V50	Tumor protein HDCMB21P	NP_003286	0.69	11
C49	Bax	NP_001179	0.76	22
U15	Toll-like receptor 3	NP_003256	0.781	4
Y01	unknown protein from clone RP11-517O1 on chrom. X	AL355476	1.03	1
W06	Voltage dependent anion channel 3	NP_005653	1.12	1
V68	Novel protein B	--	1.16	8
T25	Novel protein D	--	1.61	3

A comparison of K_d values of the Bcl-X_L-binding polypeptides (Table 2) to their frequency in the pool (Table 1) showed a 65% overlap; of the twenty lowest Bcl-X_L-binding polypeptide K_d values, thirteen were found within the top twenty most abundant Bcl-X_L-binding polypeptides, indicating a correlation between K_d and frequency. Five of the Bcl-X_L-binding polypeptides from the group with the twenty lowest K_d values, however, were observed only a single time, emphasizing the importance of post-selection characterization. Thus, the final representation of any given polypeptide within the selected pool may be determined by a number of factors: its abundance within the initial mRNA population used to prepare the library; the sum of efficiencies at each step in the mRNA display process (PCR, transcription, translation, fusion, etc.); and its affinity to the target.

As the target used in this selection was a GST fusion protein of Bcl-X_L, the specificity of each selected polypeptide was also tested by binding it to immobilized GST. The vast majority of Bcl-X_L-binding polypeptides exhibited

background levels of binding (less than 2%) to GST. Of the eight proteins that bound more than 2% to GST, five bound eight to ten fold higher to the Bcl-X_L-GST fusion protein and so were deemed specific. The three remaining proteins bound poorly to the Bcl-X_L-GST fusion relative to GST alone and so were deemed non-specific.

Many Bcl-X_L-binding polypeptides bind to the BH3 domain of Bcl-X_L

As described above, the Bcl-2 family of proteins has been shown to form homo- and hetero- dimers through the binding of the BH3 domain of one protein in the corresponding binding pocket on its partner. Only three of the selected proteins (Bim, Bak, and Bax) were previously known to contain a BH3 domain. In order to determine if the other proteins bound to the BH3 domain binding site on Bcl-X_L, a competition assay was performed. The Bak BH3 domain peptide used as a positive control was prepared by chemical synthesis and used to compete with individual Bcl-X_L-binding polypeptides in a Bcl-X_L binding assay. The effectiveness of this competition was demonstrated in a titration of competitor concentration (Figure 6). At a fixed concentration of immobilized Bcl-X_L, the Bak BH3 domain-containing peptide MGQVGRQLAIIGDDINRDYKDDDDKASA (SEQ ID NO: 152), also containing a FLAG-M2 epitope, was added at the indicated concentration along with a trace amount of a selected Talin fragment. After binding for 1 hour, the unbound material was removed and the bound protein was quantitated. The bound protein was assayed by scintillation counting, normalized to that bound in the absence of competitor, and plotted versus competitor concentration.

A competition assay was performed for each of the selected Bcl-X_L-binding polypeptides using 20 μM Bak BH3 competitor based on the titration shown in Figure 6. Due to poor competition with the Bcl-X_L-binding

polypeptides having the lowest K_d values (as determined above) a second competition was performed for some of these polypeptides using 100 μM competitor (Figure 3A). Each Bcl-X_L-binding polypeptide was incubated with immobilized Bcl-X_L in the presence of competitor and the amount bound was normalized to a comparable reaction without competitor (Figure 3A; see column labeled BakBH3 effect).

The Bcl-X_L-binding polypeptides were competed by the Bak BH3 domain, indicating that they probably bind at the same site on Bcl-X_L. The alternative explanation, a decrease in binding of the selected polypeptide at one site, due to a change in conformation of the target Bcl-X_L upon binding the competitor at a different site, was not tested in this assay. Only three of the selected proteins (clone x42, encoding SEQ ID NO: 35; clone t53, encoding SEQ ID NO: 25; and clone and w75, encoding SEQ ID NO: 37) were not competed at all by the BH3 domain, indicating that they may bind to a different site on Bcl-X_L.

Alignment of selected Bcl-X_L-binding polypeptides

Competition for binding with the Bak BH3 domain indicated that most of the Bcl-X_L-binding polypeptides that were selected were binding at the same site. Therefore, each of the polypeptides was examined for the presence of a BH3 domain sequence. A tentative assignment could be made for most polypeptides. The Bcl-X_L-binding polypeptides with the highest affinity (Table 2) are shown in Figure 7, aligned by sequence homology, where possible, to the known BH3 domains of Bim, Bak, and Bax. Most of the polypeptides have the hallmark periodicity of hydrophobic amino acids indicative of an amphipathic alpha helix. Additional homologies among the sequences are indicated by shading.

Additional Selection Experiments

Another selection to identify Bcl-X_L-GST fusion protein binders from mRNA display libraries prepared from tissue specific mRNAs of human bone marrow, brain, hippocampus, and thymus was initiated. Each library contained a unique 8 nucleotide source tag within the 5'UTR to allow specific amplification of an individual library. The source tags AACTCCTC (SEQ ID NO: 246), AATCTACC (SEQ ID NO: 247), AACAACAC (SEQ ID NO: 248), and AATATTCC (SEQ ID NO: 249) were used for the libraries derived from mRNA from human bone marrow, brain hippocampus, and thymus, respectively. Prior to initiating the selection, the libraries were pooled.

After five rounds of selection, each library was prepared individually after specific amplification using library specific primers and analyzed by cloning and sequencing. A total of 10 distinct sequence clusters were identified, of which 2 (Bim and Bax) were already identified in the previous selection. The unique sequences are shown in Figures 8A and 8B, and their corresponding nucleic acid sequences in Figure 8C. Sequences of three of the clones were from known polypeptides (SEQ ID NOS: 224-226), sequences of two of the clones were from hypothetical or unknown polypeptides whose nucleic acid sequences were found in the database (SEQ ID: 227 and 228), and sequences of two of the clones were unique polypeptide sequences (SEQ ID: 229 and 230). All of the selected Bcl-X_L-binding polypeptide sequences were from the coding region of the native protein.

The following selected polypeptides that interacted with the Bcl-X_L-GST fusion protein were synthesized and purified as described: SRP9 (clone AttB-Hc-6) and Bmf (clone AttB-Thy-34), which were unique to this selection and Bax (clone AttB-Hc-7) as a positive control for binding to the Bcl-X_L-GST fusion protein. The purified polypeptides were assayed for binding to GST and to the Bcl-X_L-GST fusion protien (Figure 9). Binding of Bax to the Bcl-X_L-

GST fusion protein was the most efficient (32%), followed by Bmf (6%) and SRP9 (0.65%). Binding of all three purified polypeptides to GST were very low, with binding percentages not higher than 0.25%.

High-throughput identification of protein-protein interactions

All of the procedures described above were essentially microcentrifuge tube based. Such systems are readily scalable through the use of microtiter techniques and are amenable to automation. In addition, the relatively laborious step of sequencing can be supplemented or replaced by array-based analysis of the pool, using, for example, Gene Discovery Arrays/Life Grids (Incyte Genomics, Palo Alto, CA) according to the manufacturer's instructions. These modifications to mRNA display technology enable its application to high-throughput, genome-wide identification of protein-protein interactions.

Cloning full length nucleic acid molecules encoding Bcl-X_L-binding polypeptides

Nucleic acid molecules encoding the full length polypeptide sequences of the identified Bcl-X_L-binding polypeptides can readily be cloned using standard hybridization or PCR cloning techniques and DNA from the source (as determined by the source label), for example, as described in Ausubel et al. (supra). An exemplary method for obtaining the full length polypeptide sequences employs, a standard nested PCR strategy that can be used with gene-specific (obtained from the nucleic acid sequence encoding the Bcl-X_L-binding polypeptide) and flanking adaptors from double stranded cDNA prepared from the source of the identified Bcl-X_L-binding polypeptide. In addition, 5' flanking sequence can be obtained using 5' RACE techniques known to those of skill in the art.

Synthesis of Bcl-X_L-binding polypeptides

Additional characteristics of the Bcl-X_L-binding polypeptides may be analyzed by synthesizing the polypeptides in various cell types or *in vitro* systems. The function of Bcl-X_L-binding polypeptides may then be examined under different physiological conditions. Alternatively, cell lines may be produced which over-express the nucleic acid encoding a Bcl-X_L-binding polypeptide, allowing purification of a Bcl-X_L-binding polypeptide for biochemical characterization, large-scale production, antibody production, or patient therapy.

For polypeptide expression, eukaryotic and prokaryotic expression systems may be generated in which nucleic acid sequences encoding Bcl-X_L-binding polypeptides are introduced into a plasmid or other vector, which is then used to transform living cells. Constructs in which the nucleic acid sequences are inserted in the correct orientation into an expression plasmid may be used for protein expression. Alternatively, portions of gene sequences encoding the Bcl-X_L-binding polypeptide, including wild-type or mutant Bcl-X_L-binding polypeptide sequences, may be inserted. Prokaryotic and eukaryotic expression systems allow various important functional domains of the Bcl-X_L-binding polypeptides to be recovered, if desired, as fusion proteins, and then used for binding, structural, and functional studies and also for the generation of appropriate antibodies. If Bcl-X_L-binding polypeptide expression induces terminal differentiation in some types of cells, it may be desirable to express the protein under the control of an inducible promoter in those cells.

Standard expression vectors contain promoters that direct the synthesis of large amounts of mRNA corresponding to the inserted nucleic acid encoding a Bcl-X_L-binding polypeptide in the plasmid-bearing cells. They may also include eukaryotic or prokaryotic origin of replication sequences allowing for their autonomous replication within the host organism, sequences

that encode genetic traits that allow vector-containing cells to be selected in the presence of otherwise toxic drugs, and sequences that increase the efficiency with which the synthesized mRNA is translated. Stable long-term vectors may be maintained as freely replicating entities by using regulatory elements of, for example, viruses (e.g., the OriP sequences from the Epstein Barr Virus genome). Cell lines may also be produced that have integrated the vector into the genomic DNA, and in this manner the gene product is produced on a continuous basis.

Expression of foreign sequences in bacteria such as *Escherichia coli* requires the insertion of the nucleic acid sequence encoding a Bcl-X_L-binding polypeptide into a bacterial expression vector. Such plasmid vectors contain several elements required for the propagation of the plasmid in bacteria, and for expression of the DNA inserted into the plasmid. Propagation of only plasmid-bearing bacteria is achieved by introducing, into the plasmid, selectable marker-encoding sequences that allow plasmid-bearing bacteria to grow in the presence of otherwise toxic drugs. The plasmid also contains a transcriptional promoter capable of producing large amounts of mRNA from the cloned gene. Such promoters may be (but are not necessarily) inducible promoters that initiate transcription upon induction. The plasmid also preferably contains a polylinker to simplify insertion of the gene in the correct orientation within the vector.

Once the appropriate expression vectors containing a nucleic acid sequence encoding a Bcl-X_L-binding polypeptide, or fragment, fusion, or mutant thereof, are constructed, they are introduced into an appropriate host cell by transformation techniques, including calcium phosphate transfection, DEAE-dextran transfection, electroporation, microinjection, protoplast fusion, and liposome-mediated transfection. The host cells that are transfected with the vectors of this invention may include (but are not limited to) *E. coli* or other

bacteria, yeast, fungi, insect cells (using, for example, baculoviral vectors for expression), or cells derived from mice, humans, or other animals. Mammalian cells can also be used to express the Bcl-X_L-binding polypeptides using, for example, a vaccinia virus expression system described, for example, in Ausubel et al. (supra).

Expression of Bcl-X_L-binding polypeptides, fusions, polypeptide fragments, or mutants encoded by cloned DNA is also possible using, for example, the T7 late-promoter expression system. This system depends on the regulated expression of T7 RNA polymerase, an enzyme encoded in the DNA of bacteriophage T7. The T7 RNA polymerase initiates transcription at a specific 23-bp promoter sequence called the T7 late promoter. Copies of the T7 late promoter are located at several sites on the T7 genome, but none is present in *E. coli* chromosomal DNA. As a result, in T7-infected *E. coli* cells, T7 RNA polymerase catalyzes transcription of viral genes but not of *E. coli* genes. In this expression system, recombinant *E. coli* cells are first engineered to carry the gene encoding T7 RNA polymerase next to the *lac* promoter. In the presence of IPTG, these cells transcribe the T7 polymerase gene at a high rate and synthesize abundant amounts of T7 RNA polymerase. These cells are then transformed with plasmid vectors that carry a copy of the T7 late promoter protein. When IPTG is added to the culture medium containing these transformed *E. coli* cells, large amounts of T7 RNA polymerase are produced. The polymerase then binds to the T7 late promoter on the plasmid expression vectors, catalyzing transcription of the inserted cDNA at a high rate. Since each *E. coli* cell contains many copies of the expression vector, large amounts of mRNA corresponding to the cloned cDNA can be produced in this system. The resulting protein can be radioactively labeled. Plasmid vectors containing late promoters and the corresponding RNA polymerases from related bacteriophages such as T3, T5, and SP6 may also be used for production of

proteins from cloned DNA. *E. coli* can also be used for expression using an M13 phage such as mGPI-2. Furthermore, vectors that contain phage lambda regulatory sequences, or vectors that direct the expression of fusion proteins, for example, a maltose-binding protein fusion protein or a glutathione-S-transferase fusion protein, also may be used for expression in *E. coli*.

Eukaryotic expression systems are useful for obtaining appropriate post-translational modification of expressed polypeptides. Transient transfection of a eukaryotic expression plasmid allows the transient production of Bcl-X_L-binding polypeptides by a transfected host cell. Bcl-X_L-binding polypeptides may also be produced by a stably-transfected mammalian cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public (e.g., see Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, Supp. 1987), as are methods for constructing such cell lines (see e.g., Ausubel et al., supra). In one example, a nucleic acid molecule encoding a Bcl-X_L-binding polypeptide, fusion, mutant, or polypeptide fragment is cloned into an expression vector that includes the dihydrofolate reductase (DHFR) gene. Integration of the plasmid and, therefore, integration of the nucleic acid sequence encoding the Bcl-X_L-binding polypeptide into the host cell chromosome is selected for by inclusion of 0.01-300 μM methotrexate in the cell culture medium (as described, for example in Ausubel et al., supra). This dominant selection can be accomplished in most cell types. Recombinant protein expression can be increased by DHFR-mediated amplification of the transfected gene. Methods for selecting cell lines bearing gene amplifications are described in Ausubel et al. (supra). These methods generally involve extended culture in medium containing gradually increasing levels of methotrexate. The most commonly used DHFR-containing expression vectors are pCVSEII-DHFR and pAdD26SV(A) (described, for example, in Ausubel et al., supra). The host cells described above or,

preferably, a DHFR-deficient CHO cell line (e.g., CHO DHFR$^-$ cells, ATCC Accession No. CRL 9096) are among those most preferred for DHFR selection of a stably-transfected cell line or DHFR-mediated gene amplification.

Eukaryotic cell expression of Bcl-X_L-binding polypeptides facilitates studies of the gene and gene products encoding Bcl-X_L-binding polypeptides, including determination of proper expression and post-translational modifications for biological activity, identifying regulatory elements located in the 5', 3', and intron regions of nucleic acid molecules encoding Bcl-X_L-binding polypeptides and their roles in tissue regulation of Bcl-X_L-binding polypeptide expression. It also permits the production of large amounts of normal and mutant proteins for isolation and purification, and the use of cells expressing Bcl-X_L-binding polypeptides as a functional assay system for antibodies generated against the protein. Eukaryotic cells expressing Bcl-X_L-binding polypeptides may also be used to test the effectiveness of pharmacological agents on apoptotic diseases or as means by which to study Bcl-X_L-binding polypeptides as components of a transcriptional activation system. Expression of Bcl-X_L-binding polypeptides, fusions, mutants, and polypeptide fragments in eukaryotic cells also enables the study of the function of the normal complete polypeptide, specific portions of the polypeptide, or of naturally occurring polymorphisms and artificially-produced mutated polypeptides. The DNA sequences encoding Bcl-X_L-binding polypeptides can be altered using procedures known in the art, such as restriction endonuclease digestion, DNA polymerase fill-in, exonuclease deletion, terminal deoxynucleotide transferase extension, ligation of synthetic or cloned DNA sequences, and site-directed sequence alteration using specific oligonucleotides together with PCR.

Another preferred eukaryotic expression system is the baculovirus system using, for example, the vector pBacPAK9, which is available from Clontech (Palo Alto, CA). If desired, this system may be used in conjunction with other protein expression techniques, for example, the myc tag approach described by Evan et al. (Mol. Cell Biol. 5:3610-3616, 1985).

Once the recombinant protein is expressed, it can be isolated from the expressing cells by cell lysis followed by protein purification techniques, such as affinity chromatography. In this example, an anti-Bcl-X_L-binding polypeptide antibody, which may be produced by the methods described herein, can be attached to a column and used to isolate the recombinant Bcl-X_L-binding polypeptides. Lysis and fractionation of Bcl-X_L-binding polypeptide-harboring cells prior to affinity chromatography may be performed by standard methods (see e.g., Ausubel et al. (supra). Once isolated, the recombinant protein can, if desired, be purified further, e.g., by high performance liquid chromatography (HPLC; e.g., see Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, Work and Burdon, Eds., Elsevier, 1980).

Polypeptides of the invention, particularly short Bcl-X_L-binding fragments, can also be produced by chemical synthesis (e.g., by the methods described in *Solid Phase Peptide Synthesis*, 2nd ed., 1984, The Pierce Chemical Co., Rockford, IL). These general techniques of polypeptide expression and purification can also be used to produce and isolate useful Bcl-X_L-binding polypeptide fragments or analogs, as described herein.

Those skilled in the art of molecular biology will understand that a wide variety of expression systems may be used to produce the recombinant Bcl-X_L-binding polypeptides. The precise host cell used is not critical to the invention. The Bcl-X_L-binding polypeptides may be produced in a prokaryotic host (e.g., *E. coli*) or in a eukaryotic host (e.g., *S. cerevisiae*, insect cells such

as Sf9 cells, or mammalian cells such as COS-1, NIH 3T3, or HeLa cells). These cells are commercially available from, for example, the American Type Culture Collection, Rockville, MD (see also Ausubel et al., supra). The method of transformation and the choice of expression vehicle (e.g., expression vector) will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel et al. (supra) and expression vehicles may be chosen from those provided, e.g., in Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, Supp. 1987.

In addition, prokaryotic and eukaryotic *in vitro* systems can be utilized for the generation of Bcl-X_L-binding polypeptides. Such methods are described, for example by Ausubel et al.(supra). Proteins can be synthesized using, for example, *in vitro* transcription and translation methods. Rabbit reticulocyte lysates, wheat germ lysates, or *E. coli* lysates can be used to translate exogenous mRNAs from a variety f eukaryotic and prokaryotic sources. Kits for the *in vitro* production of polypeptides are available, for example, from Ambion (Austin, TX).

Bcl-X_L-binding polypeptide fragments

Polypeptide fragments that incorporate various portions of Bcl-X_L-binding polypeptides are useful in identifying the domains or amino acids important for the biological activities of Bcl-X_L-binding polypeptides, and the present invention helps to identify these critical domains (Figure 4). Methods for generating such fragments are well known in the art (see, for example, Ausubel et al. (supra)) using the nucleotide sequences provided herein. For example, a Bcl-X_L-binding polypeptide fragment may be generated by PCR amplifying the desired fragment using oligonucleotide primers designed based upon the nucleic acid sequences encoding Bcl-X_L-binding polypeptides. Preferably, the oligonucleotide primers include unique restriction enzyme sites

that facilitate insertion of the fragment into the cloning site of a mammalian expression vector. This vector may then be introduced into a mammalian cell by artifice by the various techniques known in the art and described herein, resulting in the production of a Bcl-X_L-binding polypeptide gene fragment.

Bcl-X_L-binding polypeptide fragments will be useful in evaluating the portions of the polypeptide involved in important biological activities, such as protein-protein interactions. These fragments may be used alone, or as chimeric fusion proteins. Bcl-X_L-binding polypeptide fragments may also be used to raise antibodies specific for various regions of Bcl-X_L-binding polypeptides. Any portion of the Bcl-X_L-binding polypeptide amino acid sequence may be used to generate antibodies.

Bcl-X_L-binding polypeptide antibodies

In order to prepare polyclonal antibodies, Bcl-X_L-binding polypeptides, fragments of Bcl-X_L-binding polypeptides, or fusion polypeptides containing defined portions of Bcl-X_L-binding polypeptides may be synthesized in bacteria by expression of corresponding DNA sequences in a suitable cloning vehicle. Fusion proteins are commonly used as a source of antigen for producing antibodies. Two widely used expression systems for *E. coli* are *lacZ* fusions using the pUR series of vectors and *trpE* fusions using the pATH vectors. The proteins can be purified, and then coupled to a carrier protein and mixed with Freund's adjuvant (to enhance stimulation of the antigenic response in an innoculated animal) and injected into rabbits or other laboratory animals. Alternatively, protein can be isolated from Bcl-X_L-binding polypeptide-expressing cultured cells. Following booster injections at bi-weekly intervals, the rabbits or other laboratory animals are then bled and the sera isolated. The sera can be used directly or can be purified prior to use by various methods, including affinity chromatography employing reagents such

as Protein A-Sepharose, antigen-Sepharose, and anti-mouse-Ig-Sepharose. The sera can then be used to probe protein extracts from Bcl-X_L-binding polypeptide-expressing tissue electrophoretically fractionated on a polyacrylamide gel to identify Bcl-X_L-binding polypeptides. Alternatively, synthetic peptides can be made that correspond to the antigenic portions of the protein and used to innoculate the animals.

In order to generate a peptide for use in making, for example, Bcl-X_L-binding polypeptide-specific antibodies, a Bcl-X_L-binding polypeptide sequence may be expressed as a C-terminal fusion with glutathione S-transferase (GST; Smith et al., *Gene* 67:31-40, 1988). The fusion protein may be purified on glutathione-Sepharose beads, eluted with glutathione, cleaved with thrombin (at the engineered cleavage site), and purified to the degree required to successfully immunize rabbits. Primary immunizations may be carried out with Freund's complete adjuvant and subsequent immunizations performed with Freund's incomplete adjuvant. Antibody titers are monitored by Western blot and immunoprecipitation analyses using the thrombin-cleaved Bcl-X_L-binding polypeptide fragment of the Bcl-X_L-binding-GST fusion polypeptide. Immune sera are affinity purified using CNBr-Sepharose-coupled Bcl-X_L-binding polypeptide. Antiserum specificity may be determined using a panel of unrelated GST fusion proteins.

Alternatively, monoclonal Bcl-X_L-binding polypeptide antibodies may also be produced by using, as an antigen, a Bcl-X_L-binding polypeptide isolated from Bcl-X_L-binding polypeptide-expressing cultured cells or Bcl-X_L-binding polypeptide isolated from tissues. The cell extracts, or recombinant protein extracts containing Bcl-X_L-binding polypeptide, may, for example, be injected with Freund's adjuvant into mice. Several days after being injected, the mouse spleens are removed, the tissues are disaggregated, and the spleen cells are suspended in phosphate buffered saline (PBS). The spleen cells serve

as a source of lymphocytes, some of which are producing antibody of the appropriate specificity. These are then fused with permanently growing myeloma partner cells, and the products of the fusion are plated into a number of tissue culture wells in the presence of a selective agent such as hypoxanthine, aminopterine, and thymidine (HAT). The wells are then screened by ELISA to identify those containing cells making antibody capable of binding a Bcl-X_L-binding polypeptide or polypeptide fragment or mutant thereof. These are then re-plated and after a period of growth, these wells are again screened to identify antibody-producing cells. Several cloning procedures are carried out until over 90% of the wells contain single clones that are positive for antibody production. From this procedure a stable line of clones that produce the antibody is established. The monoclonal antibody can then be purified by affinity chromatography using Protein A Sepharose, ion-exchange chromatography, as well as variations and combinations of these techniques. Truncated versions of monoclonal antibodies may also be produced by recombinant methods in which plasmids are generated that express the desired monoclonal antibody fragment(s) in a suitable host.

As an alternate or adjunct immunogen to GST fusion proteins, peptides corresponding to relatively unique hydrophilic regions of Bcl-X_L-binding polypeptide may be generated and coupled to keyhole limpet hemocyanin (KLH) through an introduced C-terminal lysine. Antiserum to each of these peptides is similarly affinity-purified on peptides conjugated to BSA, and specificity is tested by ELISA and Western blotting using peptide conjugates, and by Western blotting and immunoprecipitation using Bcl-X_L-binding polypeptide, for example, expressed as a GST fusion protein.

Alternatively, monoclonal antibodies may be prepared using the Bcl-X_L-binding polypeptides described above and standard hybridoma technology (see, e.g., Kohler et al., Nature 256:495, 1975; Kohler et al., Eur. J. Immunol.

6:511, 1976; Kohler et al., Eur. J. Immunol. 6:292, 1976; Hammerling et al., In *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, New York, NY, 1981; and Ausubel et al. (supra)). Once produced, monoclonal antibodies are also tested for specific Bcl-X_L-binding polypeptide recognition by Western blot or immunoprecipitation analysis (by the methods described in Ausubel et al., supra).

Monoclonal and polyclonal antibodies that specifically recognize a Bcl-X_L-binding polypeptide (or fragments thereof), such as those described herein, are considered useful in the invention. Antibodies that inhibit the activity of a Bcl-X_L-binding polypeptide described herein may be especially useful in preventing or slowing the development of a disease caused by inappropriate expression of a wild type or mutant Bcl-X_L-binding polypeptide.

Antibodies of the invention may be produced using Bcl-X_L-binding amino acid sequences that do not reside within highly conserved regions, and that appear likely to be antigenic, as analyzed by criteria such as those provided by the Peptide Structure Program (Genetics Computer Group Sequence Analysis Package, Program Manual for the GCG Package, Version 7, 1991) using the algorithm of Jameson and Wolf (*CABIOS* 4:181, 1988). These fragments can be generated by standard techniques, e.g., by PCR, and cloned into the pGEX expression vector (Ausubel et al., supra). GST fusion proteins are expressed in *E. coli* and purified using a glutathione-agarose affinity matrix as described in Ausubel et al., supra). To generate rabbit polyclonal antibodies, and to minimize the potential for obtaining antisera that is non-specific, or exhibits low-affinity binding to a Bcl-X_L-binding polypeptide, two or three fusions are generated for each protein, and each fusion is injected into at least two rabbits. Antisera are raised by injections in series, preferably including at least three booster injections.

In addition to intact monoclonal and polyclonal anti-Bcl-X_L-binding polypeptide antibodies, the invention features various genetically engineered antibodies, humanized antibodies, and antibody fragments, including F(ab')2, Fab', Fab, Fv, and sFv fragments. Antibodies can be humanized by methods known in the art, e.g., monoclonal antibodies with a desired binding specificity can be commercially humanized (Scotgene, Scotland; Oxford Molecular, Palo Alto, CA). Fully human antibodies, such as those expressed in transgenic animals, are also features of the invention (Green et al., Nature Genetics 7:13-21, 1994).

Ladner (U.S. Patent No. 4,946,778 and 4,704,692) describes methods for preparing single polypeptide chain antibodies. Ward et al. (Nature 341:544-546, 1989) describe the preparation of heavy chain variable domains, which they term "single domain antibodies," that have high antigen-binding affinities. McCafferty et al. (Nature 348:552-554, 1990) show that complete antibody V domains can be displayed on the surface of fd bacteriophage, that the phage bind specifically to antigen, and that rare phage (one in a million) can be isolated after affinity chromatography. Boss et al. (U.S. Patent No. 4,816,397) describe various methods for producing immunoglobulins, and immunologically functional fragments thereof, which include at least the variable domains of the heavy and light chain in a single host cell. Cabilly et al. (U.S. Patent No. 4,816,567) describe methods for preparing chimeric antibodies.

Affinity reagents or polypeptides from randomized polypeptide libraries that bind tightly to a desired polypeptides, for example, Bcl-X_L-binding polypeptides, fragments of Bcl-X_L-binding polypeptides, or fusion polypeptides containing defined portions of Bcl-X_L-binding polypeptides can also be obtained, using methods known to one skilled in the art. In addition, polypeptide affinity scaffolds may be used to bind a polypeptide of interest or

to identify or optimize a polypeptide that binds to a polypeptide of interest, for example, Bcl-X_L-binding polypeptides, fragments of Bcl-X_L-binding polypeptides, or fusion polypeptides containing defined portions of Bcl-X_L-binding polypeptides. Such methods are described for example by Lipovsek et al. (WO 00/34784), hereby incorporated by reference.

Identification of additional Bcl-X_L-binding polypeptide genes

Standard techniques, such as the polymerase chain reaction (PCR) and DNA hybridization, may be used to clone Bcl-X_L-binding polypeptide homologues in other species and Bcl-X_L-binding polypeptide-related genes in humans. Bcl-X_L-binding-polypeptide-related genes and homologues may be readily identified using low-stringency DNA hybridization or low-stringency PCR with human Bcl-X_L-binding polypeptide probes or primers. Degenerate primers encoding human Bcl-X_L-binding polypeptides or human Bcl-X_L-binding polypeptide-related amino acid sequences may be used to clone additional Bcl-X_L-binding polypeptide-related genes and homologues by RT-PCR.

Alternatively, additional Bcl-X_L-binding polypeptides can be identified by utilizing consensus sequence information for Bcl-X_L-binding polypeptides to search for similar polypeptides. For example, polypeptide databases can be searched for proteins with the amphipathic alpha helix motif described above in Example 7. Candidate polypeptides containing such a motif can then be tested for their Bcl-X_L-binding properties, using methods described herein.

Assays for compounds that modulate Bcl-X_L-binding polypeptide biological activity

Bcl-X_L-binding polypeptide biological activity may be modulated in a number of different ways. For example, cellular concentrations of Bcl-X_L-binding polypeptides of can be altered, which would, in turn, affect overall Bcl-X_L-binding polypeptide biological activity. This is achieved, for example, by administering to a cell a compound that alters the concentration and/or activity of a Bcl-X_L-binding polypeptide.

We have shown herein that a number of polypeptides bind a Bcl-X_L polypeptide. Accordingly, compounds that modulate Bcl-X_L-binding polypeptide biological activity may be identified using any of the methods, described herein (or any analogous method known in the art), for measuring protein-protein interactions involving a Bcl-X_L-binding polypeptide. For example, the Bcl-X_L/Bcl-X_L-binding polypeptide assays described above may be used to determine whether the addition of a test compound increases or decreases binding activity of any (wild-type or mutant) Bcl-X_L-binding polypeptide to Bcl-X_L. A compound that increases or decreases the binding activity of a mutant Bcl-X_L-binding polypeptide may be useful for treating a Bcl-X_L-binding polypeptide-related disease, such as an apoptotic or proliferative disease. A compound that modulates Bcl-X_L-binding polypeptide biological activity may act by binding to either a Bcl-X_L-binding polypeptide or to Bcl-X_L itself, thereby reducing or preventing the biological activity of the Bcl-X_L-binding polypeptide.

Levels of Bcl-X_L-binding polypeptide may be modulated by modulating transcription, translation, or mRNA or protein turnover; such modulation may be detected using known methods for measuring mRNA and protein levels, e.g., RT-PCR and ELISA.

Test Compounds

In general, drugs for modulation of Bcl-X_L-binding polypeptide biological activity may be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their Bcl-X_L-binding polypeptide-modulatory activities should be employed whenever possible.

When a crude extract is found to modulate (i.e., stimulate or inhibit) Bcl-X_L-binding polypeptide biological activity, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having an activity that stimulates or inhibits Bcl-X_L-binding polypeptide biological activity. The same assays described herein for the detection of activities in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art. Compounds identified as being of therapeutic value may be subsequently analyzed using animal models for diseases in which it is desirable to increase or decrease Bcl-X_L-binding polypeptide biological activity.

Construction of transgenic animals and knockout animals

Characterization of Bcl-X_L-binding polypeptide genes provides information that allows Bcl-X_L-binding polypeptide knockout animal models to be developed by homologous recombination. Similarly, animal models of Bcl-X_L-binding polypeptide overproduction may be generated by integrating one or more Bcl-X_L-binding polypeptide sequences into the genome, according to

standard transgenic techniques. Moreover, the effect of Bcl-X_L-binding polypeptide gene mutations (e.g., dominant gene mutations) may be studied using transgenic mice carrying mutated Bcl-X_L-binding polypeptide transgenes or by introducing such mutations into the endogenous Bcl-X_L-binding polypeptide gene, using standard homologous recombination techniques.

Bcl-X_L-binding polypeptide knockout mice provide a tool for studying the role of Bcl-X_L-binding polypeptide in embryonic development and in disease. Moreover, such mice provide the means, *in vivo*, for testing therapeutic compounds for amelioration of diseases or conditions involving a Bcl-X_L-binding polypeptide-dependent or Bcl-X_L-binding polypeptide-affected pathway.

Construction of polypeptide knockout or overexpressing cell lines

Characterization of Bcl-X_L-binding polypeptide genes also allows Bcl-X_L-binding polypeptide cell culture models to be developed, in which the Bcl-X_L-binding polypeptide is expressed or functions at a lower level than its wild-type counterpart cell. Such cell lines can be developed using standard antisense technologies. Similarly, cell culture models of Bcl-X_L-binding polypeptide overproduction or overactivation may be generated by integrating one or more Bcl-X_L-binding polypeptide sequences into the genome, according to standard molecular biology techniques. Moreover, the effect of Bcl-X_L-binding polypeptide gene mutations (e.g., dominant gene mutations) may be studied using cell cultures model in which the cells contain and overexpress a mutated Bcl-X_L-binding polypeptide.

Bcl-X_L-binding polypeptide knockout cells provide a tool for studying the role of Bcl-X_L-binding polypeptide in cellular events, including apoptosis. Moreover, such cell lines provide the cell culture means, for testing therapeutic compounds for modulation of the apoptototic pathway.

Compounds that modulate apoptosis in these cell models can then be tested in animal models of diseases or conditions involving the apoptotic pathway.

Other Embodiments

In other embodiments, the invention includes any polypeptide that is substantially identical to a Bcl-X_L-binding polypeptide; such homologues include other substantially pure naturally-occurring Bcl-X_L-binding polypeptides as well as natural mutants; induced mutants; DNA sequences that encode polypeptides and also hybridize to the nucleic acid sequence encoding a Bcl-X_L-binding polypeptide described herein under high stringency conditions or, less preferably under low stringency conditions (e.g., washing at 2X SSC at 40°C with a probe length of at least 40 nucleotides); and proteins specifically bound by antisera directed to a Bcl-X_L-binding polypeptide. The invention also includes chimeric polypeptides that include a Bcl-X_L-binding polypeptide portion.

The invention further includes analogs of any naturally-occurring Bcl-X_L-binding polypeptide. Analogs can differ from the naturally-occurring Bcl-X_L-binding polypeptide by amino acid sequence differences, by post-translational modifications, or by both. Analogs of the invention will generally exhibit at least 85%, more preferably, 90%, and most preferably, 95% or even 99% identity with all or part of a naturally-occurring Bcl-X_L-binding polypeptide sequence. The length of sequence comparison is at least 5 amino acid residues, preferably, at least 10 amino acid residues, and more preferably, the full length of the polypeptide sequence. Modifications include *in vivo* and *in vitro* chemical derivatization of polypeptides, e.g., acetylation, carboxylation, phosphorylation, or glycosylation; such modifications may occur during polypeptide synthesis or processing or following treatment with isolated modifying enzymes. Analogs can also differ from the naturally-

occurring Bcl-X_L-binding polypeptide by alterations in primary sequence. These include genetic variants, both natural and induced (for example, resulting from random mutagenesis by irradiation or exposure to ethanemethylsulfate or by site-specific mutagenesis as described in Sambrook, Fritsch and Maniatis, Molecular Cloning: A Laboratory Manual (2d ed.), CSH Press, 1989, or Ausubel et al., supra). Also included are cyclized peptides, molecules, and analogs that contain residues other than L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth, and follows in the scope of the appended claims.

What is claimed is:

Claims

1. A substantially pure human Bcl-X_L-binding polypeptide, said polypeptide consisting of the sequence of any of SEQ ID NOS: 4-50, 63-71, and 224-228.

2. A substantially pure human Bcl-X_L-binding polypeptide, said polypeptide comprising the sequence of any of SEQ ID NOS: 51-62, 229, and 230.

3. An isolated nucleic acid molecule encoding a polypeptide of claim 1 or 2.

4. The isolated nucleic acid of claim 3, wherein said nucleic acid molecule consists of the sequence of any of SEQ ID NOS: 156-202, 215-223, and 231-235.

5. The isolated nucleic acid of claim 3, wherein said nucleic acid molecule comprises the sequence of any of SEQ ID NOS: 203-214, 236, and 237.

6. A vector comprising the isolated nucleic acid molecule of claim 3.

7. A cell comprising the isolated nucleic acid molecule of claim 3.

8. A cell comprising the vector of claim 6.

9. A method of identifying a Bcl-X_L-binding polypeptide, said method comprising the steps of:

(a) providing a population of source labeled nucleic acid-protein fusion molecules;

(b) contacting said population of nucleic acid-protein fusion molecules with a Bcl-X_L polypeptide under conditions that allow interaction between the protein portion of a nucleic acid-protein fusion molecule of said population and said Bcl-X_L polypeptide;

(c) detecting an interaction between said protein portion and said Bcl-X_L polypeptide, thereby identifying a Bcl-X_L-binding polypeptide,

10. The method of claim 9, wherein said population of source labeled nucleic acid-protein fusion molecules is derived from more than one source.

11. The method of claim 9, wherein, in step (a), said nucleic acid-protein fusion molecules are detectably-labeled.

12. The method of claim 11, wherein, in step (b), said Bcl-X_L polypeptide is immobilized on a solid support; and wherein, in step (c), the detection of an interaction between said protein portion of a nucleic acid-protein fusion molecule and said Bcl-X_L polypeptide is carried out by detecting the labeled nucleic acid-protein fusion molecule bound to said solid support.

13. The method of claim 12, wherein said solid support is a chip or a bead.

14. A method of identifying a compound that modulates binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide, said method comprising the steps of:

(a) contacting a Bcl-X_L polypeptide with (i) a Bcl-X_L-binding polypeptide, said Bcl-X_L-binding polypeptide consisting of the sequence of any of SEQ ID NOS: 4-50, 63-71, and 224-228, and (ii) a candidate compound, under conditions that allow binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide;

(b) determining the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide, wherein an increase or decrease in the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide, relative to the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide in the absence of said candidate compound, indicates a compound that modulates the binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide.

15. A method of identifying a compound that modulates binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide, said method comprising the steps of:

(a) contacting a Bcl-X_L polypeptide with (i) a Bcl-X_L-binding polypeptide, said Bcl-X_L-binding polypeptide comprising the sequence of any of SEQ ID NOS: 51-62, 229, and 230, and (ii) a candidate compound, under conditions that allow binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide;

(b) determining the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide, wherein an increase or decrease in the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide, relative to the level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide in the absence of said candidate compound, indicates a compound that modulates the binding between a Bcl-X_L polypeptide and a Bcl-X_L-binding polypeptide.

16. The method of claim 14 or 15, wherein said Bcl-X_L-binding polypeptide is part of a nucleic acid-protein fusion molecule.

17. The method of claim 14 or 15, wherein, in step (a), said Bcl-X_L polypeptide is attached to a solid support.

18. The method of claim 17, wherein said Bcl-X_L-binding polypeptide is detectably-labeled; and, in step (b), said level of binding between said Bcl-X_L polypeptide and said Bcl-X_L-binding polypeptide is determined by measuring the amount of Bcl-X_L-binding protein that binds to said solid support.

19. The method of claim 17, wherein said solid support is a chip or a bead.

20. A method of source-labeling a nucleic acid-protein fusion molecule, said method comprising the steps of:

(a) providing an RNA molecule;

(b) generating a first cDNA strand from said RNA molecule;

(c) generating a second cDNA strand complementary to said first cDNA strand, wherein said second cDNA strand comprises a nucleic acid sequence that identifies the source of said RNA molecule;

(d) generating an RNA molecule from the double stranded cDNA molecule of step (c)

(e) attaching a peptide acceptor to said RNA molecule of step (d);

(f) *in vitro* translating said RNA to generate a source labeled nucleic acid-protein fusion molecule.

21. A source-labeled nucleic acid-protein fusion molecule, said nucleic acid portion of said fusion molecule comprising a coding sequence for said protein and a label that identifies the source of said nucleic acid portion.

22. A method of identifying the source of the nucleic acid portion of a nucleic acid-protein fusion molecule, said method comprising the steps of:

(a) providing a population of nucleic acid-protein fusion molecules, said molecules comprising a source label that identifies the source of the nucleic acid portion of said nucleic acid-protein fusion molecules; and

(b) determining the identity of said source label, thereby identifying the source of the nucleic acid portion of a nucleic acid protein fusion molecule.

23. The method of claim 22, wherein said source label is cell type-specific.

24. The method of claim 22, wherein said source label is tissue-specific.

25. The method of claim 22, wherein said source label is species-specific.

26. The method of claim 22, wherein said population of nucleic acid-protein fusion molecules contains subpopulations of nucleic acid-protein fusion molecules from a plurality of sources.

FIG. 1

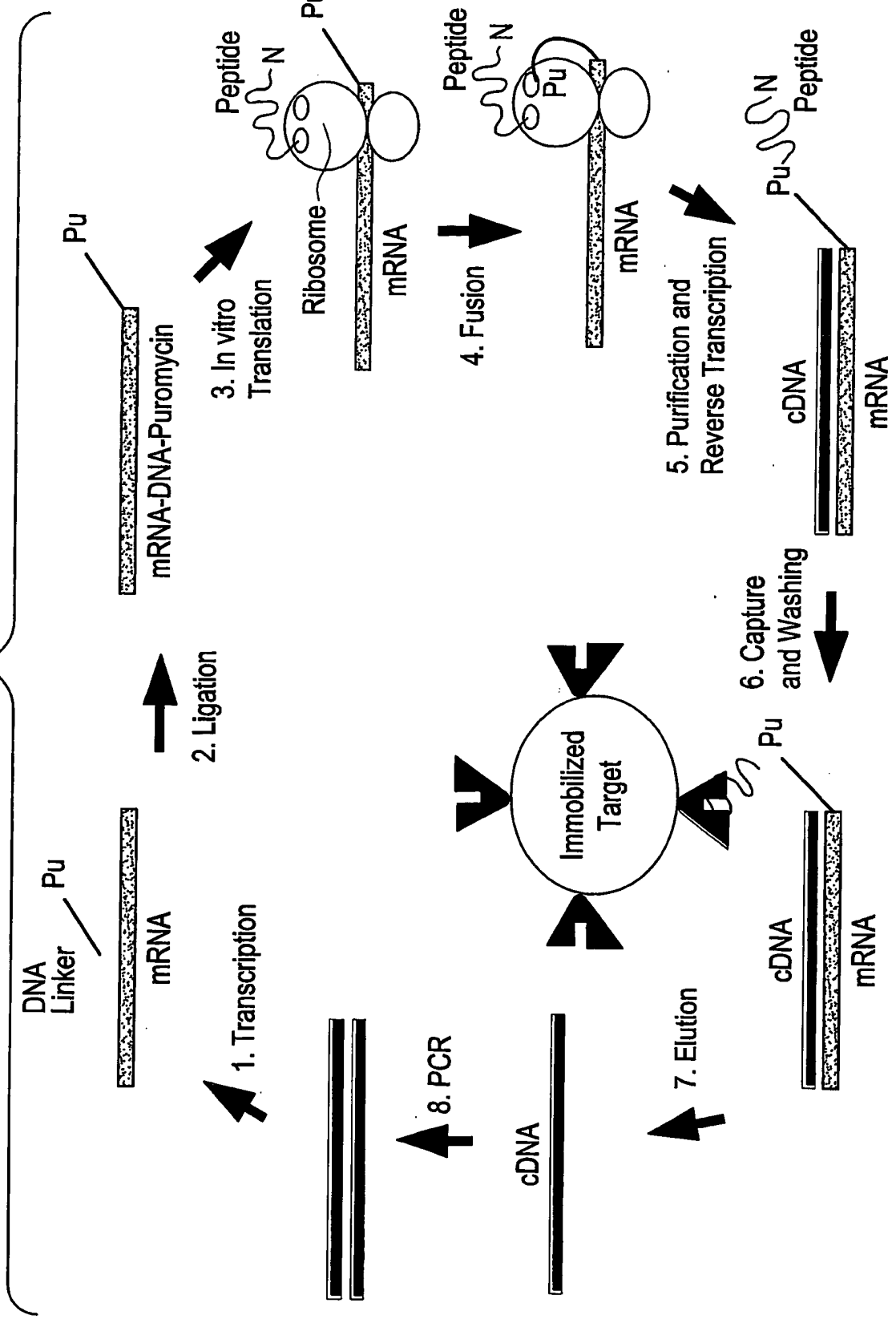

DNA Linker
Pu
mRNA
1. Transcription
2. Ligation
Pu
mRNA-DNA-Puromycin
3. In vitro Translation
Peptide
N
Pu
Ribosome
mRNA
4. Fusion
Peptide
N
Pu
mRNA
5. Purification and Reverse Transcription
N
Peptide
Pu
cDNA
mRNA
6. Capture and Washing
Immobilized Target
Pu
cDNA
mRNA
7. Elution
cDNA
8. PCR

FIG. 2A

```
Bcl-2                    M P V V H L T L T T A G D D F S R R
Bax      M P Q D A S T K K L S E C L K R I G D E L D S N G
Bak                      M G Q V G R Q L A I I G D D I N R
```

FIG. 2B

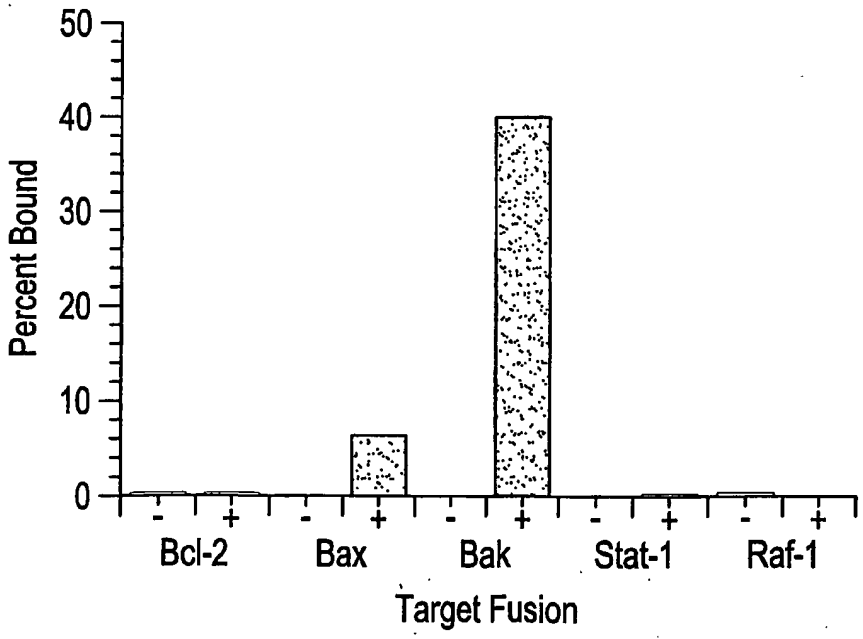


FIG. 3A-1

Number	Protein Name	Accession numbers	Clone	Protein Sequence	Kd	R value	GST Binding percent	BakBH3 [uM]	BakBH3 effect	Kidney	Liver	Marrow	Total
1	Bim	NP_006529.1	t44	ASMRQAEPADMRPEIWIAQELR RIGDEFNAYYARE	0.0019	0.97	0.3	100	0.84	42	11	36	89
2	Bak	NP_001179.1	c32	GQVGRQLAIIGDDINRRK	0.402	0.98	0.2	20	0.36	2	1	0	3
3	Bax	NP_004315.1	c49	KLSECLKRIGDELDSNMELQRMI AAVDTDSPR	0.76	0.98	1.1	20	0.06	19	0	3	22
4	Bcl2 L12	AF289220.1	y75	TGKEAILRRLVALLEEEAEVINQK LASDPALRSKLVRLSSDSFAHL	0.42	0.99	0.4	-	ND	0	2	3	5
5	Neutrophil cytosolic factor 2	NP_000424.1	t95	QRGMLYYQTEKYDLAIKDLKEAL IQLRGNN	0.00416	0.92	0.3	20	0.66	0	0	2	2
6	Talin (splice variant)	NP_006280.1	x56	GGESDTDPHFQDALMQLAKAVA SAAAALVLKAKSVAQR	0.093	0.98	0.8	20	0.21	4	0	2	6
7	Golgi SNAP receptor complex member 1	NP_004862.1	v06	GTRQDRMFETMAIEIEQLLARLT GVNDKMAEYTNA	0.467	0.94	0.1	20	0.12	0	1	0	1
8	HSCP300	AF161418	c68	AVQEDPVQREIHQDWANREYIEI ITSSIKKIAD	0.58	0.97	0.8	20	0.21	9	15	11	35
9	Syntaxin 4A	NP_004595.1	u58	ATRQALNEISARHSGIQQLERSI RELHDIFTFL	0.64	0.88	0.4	20	0.14	8	4	0	12
10	Tumor protein HDCMB21P	NP_003286.1	v50	MFSDIYGIREIADGLCLEVEGKM VSRPE	0.69	0.98	1.3	20	0.18	1	5	5	11
11	Toll-like receptor 3	NP_003256.1	u15	FWLEERDFEAGVFELEAIVNSIK RS	0.781	0.98	0.3	20	0.07	4	0	0	4
12	Voltage dependent anion channel 3	NP_005653.1	w06	MKWDTDNTLGTEISWENKLAEG LKLTLDTIFVHHVLHAPH	1.12	0.95	0.5	20	0.32	0	0	1	1
13	Aldehyde dehydrogenase	NP_003739.1	y02	RGAVFSQDKDVVQEATKVLRNA ADNFYINDR	2.5	0.99	0.1	20	0.1	0	1	1	2
14	Human retrotransposon L1	P08547	u57	TGTGAPRFIKEVQELNSALHQSD LIDIYRTLHP	2.75	0.99	0.2	20	0.08	3	0	0	3
15	TPR, nuclear pore complex-associated protein	NP_003283.1	c55	SNELTRAVEELHKLLKEARE	4.25	0.97	1.5	20	0.16	23	0	0	23

FIG. 3A-2

16	TRAP100 Thyroid hormone receptor-associated protein	AAC39855.1	u25	TYWNLLPPKRPIKEVLTDIFAKVLEKGWVDSRS	4.46	0.97	0.4	20	0.09	5	0	0	5
17	Parathyroid hormone receptor	NM_000316.1	c56	LFTILLTLWTMRCSSTPSG	5.24	0.88	0.3	20	0.19	1	0	0	1
18	Calpain	P07384	v53	AGEDMEISVKELRTILNRIISKHKDLRT	7.5	0.92	0.4	20	0.08	1	3	0	4
19	Occludin, tight junction protein	NP_002529.1	x33	GLREESEEYMAAADEYNRLKQVKQPA	8.9	0.89	0.2	100	0	1	0	0	1
20	Human nGAP protein	NP_004832.1	c26	KGIISRLMSVEEELKRDHAEMQAGCGLQTEDHLMPRRSAFASLDAVNARLMSALTPAXRYVXHCXPL	9.6	0.91	0.2	20	0.02	0	1	0	1
21	Ankryn 2	NP_001139.2	x69	WERIEERLAYIADHLGFSWTELARAL	*		3.7	20	0.5	3	0	0	3
22	Sterol regulatory element binding txn factor (splice variant)	NP_004167.2	w17	ARGDFAQAAQQLWLALRALGRPLPTSH	PD	-	0.4	20	0.27	0	0	5	5
23	Diacylglycerol kinase Zeta	NP_003637.1	y49	GSSKDLAKHIQVVCDGMDLTPKIHDLKPQC	PD	-	0.5	20	0.31	0	2	0	2
24	Translin	NP_004613.1	c08	GFLAAEQDIREEIRKVVQSLEQTAREVLTLLQG	PD	-	0.6	20	0.21	1	0	0	1
25	Hep C associated,interferon-	NP_006408.1	t53	LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG	PD	-	1	20	1	1	0	1	2
26	Mitochondrial NADH dehydrogenase subunit 1	NP_008342.1	c24	ANLLLLMVPILIAMAFLMLTERKILGYIQPR	5.04	0.95	1.4	20	0.79	1	0	1	2
27	Mitochondrial ATP synthase F0 subunit 8	NP_008346.1	c57	LRLNTTVWPTIITPILLTLFLITNRLITTR	PD	-	1.4	20	0.77	2	0	0	2
28	Mitochondrial NADH dehydrogenase chain 5	NP_008352.1	c46	TLYLKLTALAVTFLGLLTALDLNYPT	PD	-	1.5	20	0.39	1	0	0	1
29	Hypothetical protein DFKZp434e171	T14795	x68	AGVFSAEPSPFPQTRRSMVFARHLREVGDEFRSRHLNSTDDADE	4.95	0.92	0.6	20	0.21	1	0	1	2

FIG. 3A-3

30	Hypothetical protein DKFZp566f2124	AL117558.1	y16	GLKLATVAASMDRVPKVTPSSAI SSIARENHEPERLGLNGIAETT	PD	-	0.5	20	0.2	0	2	0	2
31	Hypothetical protein KIAA1501	AB040934.1	x72	MRDLPGHYYETLKFLVGHLKTIA DHR	*		11	-	ND	2	0	1	3
32	Hypothetical protein DKFZp586ho623	NM_017540.1	v47	CGGRMEDIPCSRVGHIYRKYVP YKVPAGVSLARNLKRVADWM	0.0129	0.9	0.4	20	0.36	0	1	15	16
33	unknown protein from cDNA: FLJ21691 fis, clone COL09555	AK025344.1	y37	ALSWIEMDTEMEMLLARFRRTP GDLHLDHSVHLCAHP	0.41	0.96	0.7	-	ND	0	1	0	1
34	unknown protein from Mitochondrial DNA	HSM801305	c82	TSTLPHIRRTR	1.62	0.9	0.3	20	0.19	1	0	0	1
35	Unknown protein from Homo sapiens cDNA: FLJ23277 fis, clone HEP03322	AK026930.1	x42	NGNLFASFIADS	2.04	0.99	0	20	1	12	2	2	16
36	unknown protein from Homo sapiens cDNA: FLJ22171 fis, clone HRC00654	AK025824.1	t42	ILTSPWTTSSGLWPRLQKAAEAF KQLNQP	*		3.6	20	0.76	0	0	5	5
37	unknown protein from cDNA FLJ23179 fis, clone LNG10890	AK026832.1	w75	RTLQPRLLQNQQQHLPALPIWFL LQWLRLHPL	PD	-	0.7	20	1	0	0	1	1
38	unknown protein from clone RP5-889J22 on chromosome 22q13.1	AL049850.1	x36	MAVIINELSQRDSCGPLKISLNN KILVYGNLFSSFTP	PD	-	0.7	20	0.34	1	0	0	1
39	Unknown protein from Mitochondrial DNA	AK000348	c76	GLAKKSKRNPANLTPP	PD	-	1.1	20	0.21	1	0	0	1
40	Unknown protein from Homo sapiens chromosome X	AF030876.1	c23	SSQALRIHQWLHLFSDFTST	PD	-	1.3	20	0.36	1	0	0	1
41	unknown protein bfrom clone RP11-141E20 on chromosome 1q31.2-31.3	AL136136	y86	GQVGRQLAIIGDDINRRK	NB		0.1	-	ND	0	1	0	1

FIG. 3A-4

42	unknown protein from chromosome 21q22.2, cosmid clone:D37D12, CBR1-HLCS region	AP000703.1	w52	GVSEAEGTFPLSTFLLGIASRLR SVA	PD	-	0	20	0.75	0	0	3	3
43	Unknown protein from Homo sapiens clone RP11-198M19, homology to retrotransposon	AC020550.4	u33	RAPRFIKQILLDLKREIDFNVRLV EYFNPLS	PD	-	0.2	20	0.78	3	0	0	3
44	Unknown protein with RNA Homology to Murine	AF014453.1	y74	IVAIIAGRLRMLGDQFNGELEAS AKN	PD	-	0.2	100	0.23	0	2	0	2
45	Unknown protein from CpG island	gi 1038837	v56	LALAYYSSRQYASALKHIAEIIER GIRQH	PD	-	0.3	20	0.19	0	1	0	1
46	unknown protein from clone 425C14 on chromosome 6q22	Z99129.1	v72	AAMLLDRRGTECDLWINEMSLL HKIVQDVYGTPHPPHS	0.28	0.99	2.5	20	0.2	0	1	0	1
47	unknown protein from Human genomic DNA of Xq28 with MTM1 and MTMR1 genes	AF002223.1	u46	PWQYKPIADLYRGRESRPSAPR	1.02	0.95	ND	-	ND	2	0	0	2
48	unknown protein from clone RP11-517O1 on chromosome X	AL355476.12	y01	LFSVLLRYLADNFLPGGS	1.03	0.88	0.7	20	0.08	0	1	0	1
49	Unknown protein from PAC clone RP5-1021I20 from14q24.3	AC005520.2	u96	DWQVLLGKLLWKIDNPGI	2.66	0.89	1	20	0.14	1	0	0	1
50	unknown protein from in DNA of chromosomes 8, 10, 14, 16	multiple matches	v66	GAMEREWAMFLRAASSRIRGGV	2.38	0.94	0.1	20	0.11	0	1	0	1
51	Sequence/protein not in database	AL359079	c39	VHNFGRHWGLPLSFLLNYPLFLS P	2.54	0.92	0.5	-	ND	3	0	0	3
52	Sequence/protein not in database	none	t08	ASMAPVGRDAETLQKQKETIKAF LKKLEALMASNDNANKT	2.23	0.98	2.7	20	0.07	0	0	1	1

FIG. 3A-5

53	Sequence/protein not in database	none	v68	CREQAELTGLRLASLGLKFNKIV HSSMTRAIET	1.16	0.94	0.9	-	ND	3	5	0	8
54	Sequence/protein not in database	none	t25	GTRISDMLKLIADTWQRNCCPA	1.61	0.92	0.8	20	0.26	0	1	2	3
55	Sequence/protein not in database	partial homologies only	x50	EQASVKYVILDMYRALLTLMNTS TAT	4.57	0.93	0.1	-	ND	1	0	0	1
56	Sequence/protein not in database	None	c54	EDLESVLIRLINWAKGSPIP	PD	-	0.4	-	ND	1	0	0	1
57	Sequence/protein not in database	none	x52	RPVSFCGAVWTLNRAIGRHFVR GSR	0.17	0.87	4.2	20	0.36	1	0	0	1
58	Sequence/protein not in database	none	y63	HAVVARLLHIGAIMFQRLDFIEQL SAPPA	PD	-	0.7	20	0.48	0	2	0	2
59	Sequence/protein not in database	none	c80	GQGTLWGSGMEAWLATVLKALP WHPTYQLEP	PD	-	1.3	-	ND	1	0	0	1
60	Sequence/protein not in database	partial homologies only	c21	IAQATKATIDKWNCIKLKIFYTSK KEAS	0.07	0.89	1.4	100	0.44	3	0	0	3
61	Sequence/protein not in database	partial homologies only	v18	VVDVPDFIVWLEEAVSDLHRAL	0.086	0.93	2.5	20	0.37	1	11	6	18
62	Sequence/protein not in database	none	v17	QRRGNEFQLRDLADAWDLSSRS RQRGWQMPNCRSRRGPG	PD	-	2.1	-	ND	0	1	0	1
	PD = Poorly determined, ND = Not Determined, NB = No Binding, * = High background binding												
	Proteins out of frame or from UTR												

FIG. 3A-6

64	Transforming growth factor (reversed orientation)	XM_003758	v84	FVRSVGWRLQNIGDDMDHAICG HDVRLG									
65	L-plastin (3' UTR)	NM_002298	t67	--SGLRKPTCGSSQR									
66	Arsenate resistance protein (reversed orientation)	NM_015908	y39	AGTQPLILAQFMRVGGDELLHFL LW									
67	K-Ras oncogene (3' UTR)	NM_004985	x91	MDTIKGFDLITNFQVVADALNISL LPNPLATA									
68	Lysosomal pepstatin insensitive protease (3' UTR)	AF039704	c1	ATWMKTLQGLLDRIQAFPSSPH									
69	MYBPC3 (3' UTR)	NM_000256	c41	EANRKQPKPNNSSTAYYNFTGV SILPSYKP									
70	cDNA FLJ20617 (not in predicted ORF)	AK000624	c4	GSLTHHINNIKPSSTR									
71	UDP glucoronosyl transferase 2B4 precursor (out of frame)	XM_003548	w90	VSCWPSYLKYPLSTASASLLATQ LKSIA									

– indicates that the primary nucleic acid sequence was truncated

FIG. 3B-1

SEQ ID NO 1: BimL (Corresponds to the nucleic acid sequence of SEQ ID NO: 153)
ASMRQAEPADMRPEIWIAQELRRIGDEFNAYYARE

SEQ ID NO: 2 Bak (Corresponds to the nucleic acid sequence of SEQ ID NO: 154)
GQVGRQLAIIGDDINRRK

SEQ ID NO: 3 Bax (Corresponds to the nucleic acid sequence of SEQ ID NO: 155)
KLSECLKRIGDELDSNMELQRMIAAVDTDSPR

SEQ ID NO: 4 Bcl2 L12 (Corresponds to the nucleic acid sequence of SEQ ID NO: 156)
TGKEAILRRLVALLEEEAEVINQKLASDPALRSKLVRLSSDSFA
HL

SEQ ID NO: 5 Neutrophil cytosolic factor 2 (Corresponds to the nucleic acid sequence of SEQ ID NO: 157)
QRGMLYYQTEKYDLAIKDLKEALIQLRGNN

SEQ ID NO: 6 Talin (splice variant) (Corresponds to the nucleic acid sequence of SEQ ID NO: 158)
GGESDTDPHFQDALMQLAKAVASAAAALVLKAKSVAQR

SEQ ID NO: 7 Golgi SNAP receptor complex member 1 (Corresponds to the nucleic acid sequence of SEQ ID NO: 159)
GTRQDRMFETMAIEIEQLLARLTGVNDKMAEYTNA

SEQ ID NO: 8 HSCP300 (Corresponds to the nucleic acid sequence of SEQ ID NO: 160)
AVQEDPVQREIHQDWANREYIEIITSSIKKIAD

SEQ ID NO: 9 Syntaxin 4A (Corresponds to the nucleic acid sequence of SEQ ID NO: 161)
ATRQALNEISARHSGIQQLERSIRELHDIFTFL

SEQ ID NO: 10 Tumor protein HDCMB21P (Corresponds to the nucleic acid sequence of SEQ ID NO: 162)
MFSDIYGIREIADGLCLEVEGKMVSRPE

SEQ ID NO: 11 Toll-like receptor 3 (Corresponds to the nucleic acid sequence of SEQ ID NO: 163)
FWLEERDFEAGVFELEAIVNSIKRS

SEQ ID NO: 12 Voltage dependent anion channel 3 (Corresponds to the nucleic acid sequence of SEQ ID NO: 164)
MKWDTDNTLGTEISWENKLAEGLKLTLDTIFVHHVLHAPH

FIG. 3B-2

SEQ ID NO: 13 Aldehyde dehydrogenase (Corresponds to the nucleic acid sequence of SEQ ID NO: 165)
RGAVFSQDKDVVQEATKVLRNAADNFYINDR

SEQ ID NO: 14 Human retrotransposon L1 (Corresponds to the nucleic acid sequence of SEQ ID NO: 166)
TGTGAPRFIKEVQELNSALHQSDLIDIYRTLHP

SEQ ID NO: 15 TPR, nuclear pore complex-associated protein (Corresponds to the nucleic acid sequence of SEQ ID NO: 167)
SNELTRAVEELHKLLKEARE

SEQ ID NO: 16 TRAP100 Thyroid hormone receptor-associated protein (Corresponds to the nucleic acid sequence of SEQ ID NO: 168)
TYWNLLPPKRPIKEVLTDIFAKVLEKGWVDSRS

SEQ ID NO: 17 Parathyroid hormone receptor (Corresponds to the nucleic acid sequence of SEQ ID NO: 169)
LFTILLTLWTMRCSSTPSG

SEQ ID NO: 18 Calpain (Corresponds to the nucleic acid sequence of SEQ ID NO: 170)
AGEDMEISVKELRTILNRIISKHKDLRT

SEQ ID NO: 19 Occludin, tight junction protein (Corresponds to the nucleic acid sequence of SEQ ID NO: 171)
GLREESEEYMAAADEYNRLKQVKQPA

SEQ ID NO: 20 Human nGAP protein (Corresponds to the nucleic acid sequence of SEQ ID NO: 172)
KGIISRLMSVEEELKRDHAEMQAGCGLQTEDHLMPRRSAFASL
DAVNARLMSALTPAXRYVXHCXPL

SEQ ID NO: 21 Ankryn 2 (Corresponds to the nucleic acid sequence of SEQ ID NO: 173)
WERIEERLAYIADHLGFSWTELARAL

SEQ ID NO: 22 Sterol regulatory element binding txn factor (splice variant) (Corresponds to the nucleic acid sequence of SEQ ID NO: 174)
ARGDFAQAAQQLWLALRALGRPLPTSH

SEQ ID NO: 23 Diacylglycerol kinase Zeta (Corresponds to the nucleic acid sequence of SEQ ID NO: 175)
GSSKDLAKHIQVVCDGMDLTPKIHDLKPQC

FIG. 3B-3

SEQ ID NO: 24 Translin (Corresponds to the nucleic acid sequence of SEQ ID NO: 176)
GFLAAEQDIREEIRKVVQSLEQTAREVLTLLQG

SEQ ID NO: 25 Hep C associated,interferon-induced microtubular (Corresponds to the nucleic acid sequence of SEQ ID NO: 177)
LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG

SEQ ID NO: 26 Mitochondrial NADH dehydrogenase subunit 1 (Corresponds to the nucleic acid sequence of SEQ ID NO: 178)
ANLLLLMVPILIAMAFLMLTERKILGYIQPR

SEQ ID NO: 27 Mitochondrial ATP synthase F0 subunit 8 (Corresponds to the nucleic acid sequence of SEQ ID NO: 179)
LRLNTTVWPTIITPILLTLFLITNRLITTR

SEQ ID NO: 28 Mitochondrial NADH dehydrogenase chain 5 (Corresponds to the nucleic acid sequence of SEQ ID NO: 180)
TLYLKLTALAVTFLGLLTALDLNYPT

SEQ ID NO: 29 Hypothetical protein DFKZp434e171 (Corresponds to the nucleic acid sequence of SEQ ID NO: 181)
AGVFSAEPSPFPQTRRSMVFARHLREVGDEFRSRHLNSTDDAD
E

SEQ ID NO: 30 Hypothetical protein DKFZp566f2124 (Corresponds to the nucleic acid sequence of SEQ ID NO: 182)
GLKLATVAASMDRVPKVTPSSAISSIARENHEPERLGLNGIAET
T

SEQ ID NO: 31 Hypothetical protein KIAA1501 (Corresponds to the nucleic acid sequence of SEQ ID NO: 183)
MRDLPGHYYETLKFLVGHLKTIADHR

SEQ ID NO: 32 Hypothetical protein DKFZp586ho623 (Corresponds to the nucleic acid sequence of SEQ ID NO: 184)
CGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVADWM

SEQ ID NO: 33 unknown protein from cDNA: FLJ21691 fis, clone COL09555 (Corresponds to the nucleic acid sequence of SEQ ID NO: 185)
ALSWIEMDTEMEMLLARFRRTPGDLHLDHSVHLCAHP

FIG. 3B-4

SEQ ID NO: 34 unknown protein from Mitochondrial DNA (Corresponds to the nucleic acid sequence of SEQ ID NO: 186)
T S T L P H I R R T R

SEQ ID NO: 35 Unknown protein from Homo sapiens cDNA: FLJ23277 fis, clone HEP03322 (Corresponds to the nucleic acid sequence of SEQ ID NO: 187)
N G N L F A S F I A D S

SEQ ID NO: 36 unknown protein from Homo sapiens cDNA: FLJ22171 fis, clone HRC00654 (Corresponds to the nucleic acid sequence of SEQ ID NO: 188)
I L T S P W T T S S G L W P R L Q K A A E A F K Q L N Q P

SEQ ID NO: 37 unknown protein from cDNA FLJ23179 fis, clone LNG10890 (Corresponds to the nucleic acid sequence of SEQ ID NO: 189)
R T L Q P R L L Q N Q Q Q H L P A L P I W F L L Q W L R L H P L

SEQ ID NO: 38 unknown protein from clone RP5-889J22 on chromosome 22q13.1 (Corresponds to the nucleic acid sequence of SEQ ID NO: 190)
M A V I I N E L S Q R D S C G P L K I S L N N K I L V Y G N L F S S F T P

SEQ ID NO: 39 Unknown protein from Mitochondrial DNA (Corresponds to the nucleic acid sequence of SEQ ID NO: 191)
G L A K K S K R N P A N L T P P

SEQ ID NO: 40 Unknown protein from Homo sapiens chromosome X (Corresponds to the nucleic acid sequence of SEQ ID NO: 192)
S S Q A L R I H Q W L H L F S D F T S T

SEQ ID NO: 41 unknown protein bfrom clone RP11-141E20 on chromosome 1q31.2-31.3 (Corresponds to the nucleic acid sequence of SEQ ID NO: 193)
G Q V G R Q L A I I G D D I N R R K

SEQ ID NO: 42 unknown protein from chromosome 21q22.2, cosmid clone:D37D12, CBR1-HLCS region (Corresponds to the nucleic acid sequence of SEQ ID NO: 194)
G V S E A E G T F P L S T F L L G I A S R L R S V A

SEQ ID NO: 43 Unknown protein from Homo sapiens clone RP11-198M19, homology to retrotransposon (Corresponds to the nucleic acid sequence of SEQ ID NO: 195)
R A P R F I K Q I L L D L K R E I D F N V R L V E Y F N P L S

FIG. 3B-5

SEQ ID NO: 44 Unknown protein with RNA Homology to Murine retrovirus readthrough seq. (Corresponds to the nucleic acid sequence of SEQ ID NO: 196)
I V A I I A G R L R M L G D Q F N G E L E A S A K N

SEQ ID NO: 45 Unknown protein from CpG island (Corresponds to the nucleic acid sequence of SEQ ID NO: 197)
L A L A Y Y S S R Q Y A S A L K H I A E I I E R G I R Q H

SEQ ID NO: 46 unknown protein from clone 425C14 on chromosome 6q22 (Corresponds to the nucleic acid sequence of SEQ ID NO: 198)
A A M L L D R R G T E C D L W I N E M S L L H K I V Q D V Y G T P H P P H S

SEQ ID NO: 47 unknown protein from Human genomic DNA of Xq28 with MTM1 and MTMR1 genes (Corresponds to the nucleic acid sequence of SEQ ID NO: 199)
P W Q Y K P I A D L Y R G R E S R P S A P R

SEQ ID NO: 48 unknown protein from clone RP11-517O1 on chromosome X (Corresponds to the nucleic acid sequence of SEQ ID NO: 200)
L F S V L L R Y L A D N F L P G G S

SEQ ID NO: 49 Unknown protein from PAC clone RP5-1021I20 from14q24.3 (Corresponds to the nucleic acid sequence of SEQ ID NO: 201)
D W Q V L L G K L L W K I D N P G I

SEQ ID NO: 50 unknown protein from in DNA of chromosomes 8, 10, 14, 16 (Corresponds to the nucleic acid sequence of SEQ ID NO: 202)
G A M E R E W A M F L R A A S S R I R G G V

SEQ ID NO: 51 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 203)
V H N F G R H W G L P L S F L L N Y P L F L S P

SEQ ID NO: 52 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 204)
A S M A P V G R D A E T L Q K Q K E T I K A F L K K L E A L M A S N D N A N K T

SEQ ID NO: 53 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 205)
C R E Q A E L T G L R L A S L G L K F N K I V H S S M T R A I E T

FIG. 3B-6

SEQ ID NO: 54 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 206)
GTRISDMLKLIADTWQRNCCPA

SEQ ID NO: 55 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 207)
EQASVKYVILDMYRALLTLMNTSTAT

SEQ ID NO: 56 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 208)
EDLESVLIRLINWAKGSPIP

SEQ ID NO: 57 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 209)
RPVSFCGAVWTLNRAIGRHFVRGSR

SEQ ID NO: 58 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 210)
HAVVARLLHIGAIMFQRLDFIEQLSAPPA

SEQ ID NO: 59 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 211)
GQGTLWGSGMEAWLATVLKALPWHPTYQLEP

SEQ ID NO: 60 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 212)
IAQATKATIDKWNCIKLKIFYTSKKEAS

SEQ ID NO: 61 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 213)
VVDVPDFIVWLEEAVSDLHRAL

SEQ ID NO: 62 Sequence/protein not in database (Corresponds to the nucleic acid sequence of SEQ ID NO: 214)
QRRGNEFQLRDLADAWDLSSRSRQRGWQMPNCRSRRGPG

SEQ ID NO: 63 Proline/glutamine rich splicing factor (Corresponds to the nucleic acid sequence of SEQ ID NO: 215)
RGLWVDRVLEEWGLEPRQ

FIG. 3B-7

SEQ ID NO: 64 Transforming growth factor (Corresponds to the nucleic acid sequence of SEQ ID NO: 216)
F V R S V G W R L Q N I G D D M D H A I C G H D V R L G

SEQ ID NO: 65 L-plastin (Corresponds to the nucleic acid sequence of SEQ ID NO: 217)
- - S G L R K P T C G S S Q R

SEQ ID NO: 66 Arsenate resistance protein (Corresponds to the nucleic acid sequence of SEQ ID NO: 218)
A G T Q P L I L A Q F M R V G G D E L L H F L L W

SEQ ID NO: 67 K-Ras oncogene (Corresponds to the nucleic acid sequence of SEQ ID NO: 219)
M D T I K G F D L I T N F Q V V A D A L N I S L L P N P L A T A

SEQ ID NO: 68 Lysosomal pepstatin insensitive protease (Corresponds to the nucleic acid sequence of SEQ ID NO: 220)
A T W M K T L Q G L L D R I Q A F P S S P H

SEQ ID NO: 69 MYBPC3 (Corresponds to the nucleic acid sequence of SEQ ID NO: 221)
E A N R K Q P K P N N S S T A Y Y N F T G V S I L P S Y K P

SEQ ID NO: 70 cDNA FLJ20617 (Corresponds to the nucleic acid sequence of SEQ ID NO: 222)
G S L T H H I N N I K P S S T R

SEQ ID NO: 71 UDP glucoronosyl transferase 2B4 precursor (Corresponds to the nucleic acid sequence of SEQ ID NO: 223)
V S C W P S Y L K Y P L S T A S A S L L A T Q L K S I A

FIG. 3C-1

SEQ ID NOS:72 and 153

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTACTTCACAATGGCTTCCATGA
GGCAGGCTGAACCTGCAGATATGCGCCCAGAGATATGGATCGCCCAAGAGTTGCGGCGTATTG
GAGACGAGTTTAACGCCTACTATGCAAGGGAGGATTACAAAGACGATGACGATAAGGCATCCG
CTATTTAAAA

SEQ ID NOS:73 and 154

TACTATTTACAATTCTCCTAACACAATGGGGGCAGGTGGGGACGGCAGCTCGCCATCATCGGG
GACGACATCAACCGACGGAAAGATTACAAAGACGATGACGATAAGGCATCCGCTATTAAAAAA

SEQ ID NOS:74 and 155

TTTACAATTCTCCTAACACAATGAAGCTGAGCGAGTGTCTCAAGCGCATCGGGGACGAACTGG
ACAGTAACATGGAGCTGCAGAGGATGATTGCCGCCGTGGACACAGACTCCCCCCGAGATTACA
AAGACGATGACGATAAGGCATCCGCTATTAAAAA

SEQ ID NOS:75 and 156

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGACAGGGAAGG
AAGCCATACTGCGGAGGCTGGTGGCCCTGCTGGAGGAGGAGGCAGAAGTCATTAACCAGAAGC
TGGCCTCGGACCCCGCCCTGCGCAGCAAGCTGGTCCGCCTGTCCTCCGACTCTTTCGCCCACC
TGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:76 and 157

GACTCACTATAGGGACAATTACTATTTACAATTCTTACTTCCAACGAGGGATGCTCTACTACC
AGACAGAGAAATATGATTTGGCTATCAAAGACCTTAAAGAAGCCTTGATTCAGCTTCGAGGGA
ACAATGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:77 and 158

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGGTGGGGAAA
GTGATACTGACCCCCACTTCCAGGATGCGCTAATGCAGCTCGCCAAAGCTGTGGCAAGTGCTG
CAGCTGCCCTGGTCCTCAAGGCCAAGAGTGTGGCCCAACGAGATTACAAAGACGATGACGATA
GGGCATCCGCTATTTAAAA

SEQ ID NOS:78 and 159

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGGAACACGCC
AAGACAGAATGTTTGAGACAATGGCGATTGAGATTGAACAACTTTTGGCAAGGCTTACAGGGG
TAAATGATAAAATGGCAGAATATACCAACGCTGATTACAAAGACGATGACGATAAGGCATCCG
CTATTTAAAA

FIG. 3C-2

SEQ ID NOS:79 and 160

CTATTTACAATTCTCCTAACACAATGGCGGTACAGGAGGATCCGGTGCAGCGGGAGATTCACC
AGGACTGGGCTAACCGGGAGTACATTGAGATAATCACCAGCAGCATCAAGAAAATCGCAGACT
TTCTCAACTCGTTCGATTACAAAGACGATGACGATAAGGCATCCGCTATTAAAAA

SEQ ID NOS:80 and 161

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGCGACTCGAC
AGGCCTTAAATGAGATCTCGGCCCGGCACAGTGGGATCCAGCAGCTTGAACGCAGTATTCGTG
AGCTGCACGACATATTCACTTTTCTGGCTACCGAAGTGCGAGATTACAAAGACGATGACGATA
AGGCATCCGCTATTTAAAA

SEQ ID NOS:81 and 162

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGATGTTCTCCG
ACATCTACGGGATCCGGGAGATCGCGGACGGGTTGTGCCTGGAGGTGGAGGGGAAGATGGTCA
GTAGGCCAGAGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:82 and 163

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGTTTTGGCTGG
AAGAAAGGGACTTTGAGGCGGGTGTTTTTGAACTAGAAGCAATTGTTAACAGCATCAAAAGAA
GCGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:83 and 164

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTACTTCAATACAATGAAATGGG
ACACAGACAATACTCTAGGGACAGAAATCTCTTGGGAGAATAAGTTGGCTGAAGGGTTGAAAC
TGACTCTTGATACCATATTTGTACATCACGTCCTGCATGCCCCACACGATTACAAAGACGATG
ACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:84 and 165

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGCGGGGGGCAG
TGTTCTCCCAGGATAAGGACGTCGTGCAGGAGGCCACAAAGGTGCTGAGGAATGCTGCCGACA
ACTTCTACATCAACGACAGGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:85 and 166

GACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGACCGGTACAGGAGCAC
CCAGATTCATAAAGGAAGTCCAGGAATTGAACTCAGCTCTACATCAATCGGACCTAATAGACA
TCTACAGAACTCTCCACCCCGCTGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAA
A

FIG. 3C-3

SEQ ID NOS:86 and 167

TTTACAATTCTCCTAACACAATGACAAAGAGCAATGAACTAACCCGGGCAGTAGAGGAACTAC
ACAAACTTTTGAAAGAAGCTAGGGAAGATTACAAAGACGATGACGATAAGGCATCCGCTATTT
AAAA

SEQ ID NOS:87 and 168

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGACCTACTGGA
ACCTGCTGCCCCCCAAGCGGCCCATCAAAGAGGTGCTGACGGACATCTTTGCCAAGGTGCTGG
AGAAGGGCTGGGTGGACAGCCGCTCCATCCACGATTACAAAGACGATGACGATAAGGCATCCG
CTATTTAAAA

SEQ ID NOS:88 and 169

CTATTTACAATTCTCCTAACACTATGGACTATGAGATGCTCTTCAACTCCTTCAGGGATTACA
AAGACGATGACGATAAGGCATCCGCTATTAAAAA

SEQ ID NOS:89 and 170

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGCCGGGGAGG
ACATGGAGATCAGCGTGAAGGAGTTGCGGACAATCCTCAATAGGATCATCAGCAAACACAAAG
ACCTGCGGACCGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:90 and 171

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGGACTAAGAG
AAGAAAGTGAAGAGTACATGGCTGCTGCTGATGAATACAATAGACTGAAGCAAGTGAAGCAAC
CTGCAGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:91 and 172

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGAAGGGCATCA
TCAGCAGGTTGATGTCCGTGGAGGAAGAACTGAAGAGGGACCACGCAGAGATGCAAGCGGCTG
TGGACTCCAAACAGAAGATCATTGATGCCCAGGAGAAGCGCATTGCCTCGTTGGATGCCGCCA
ATGCCCGCCTCATGAGTGCCCTGACCCAGCTGAAAGAGAGGTACAGCATGCAAGCCCGTAACG
GCATCTCCCCCACCAACCCCGCGGATTACAAAGACGATGACGATAAGGCATCCACTATTTAAA
AAA

SEQ ID NOS:92 and 173

TAATACGACTCACTATAGGGACAAATACTATTTACAATTCTCCTAACACAATGTGGGAACGGA
TTGAGGAAAGGCTGGCTTATATTGCTGATCACCTTGGCTTCAGCTGGACAGAATTAGCAAGAG
CGCTGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

FIG. 3C-4

SEQ ID NOS:93 and 174

TAATACGACTCACTATAGGGGACAATTACTATTTACAATTGCTTACTTCACAATGGCTCGGGG
AGACTTTGCCCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGCGGGCACTGGGCCGGCCCCTGCC
CACCTCCCACGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:94 and 175

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGTGGTGGATG
TGCCAGATTTTATAGTCTGGCTTGAGGAGGCAGTATCTGATTTACATAGGGCCCTCGATTACA
AAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:95 and 176

CTTTTACAATTCTCCTAACACAATGGGCTTTTTGGCTGCCGAGCAGGACATCCGAGAGGAAAT
CAGAAAAGTTGTACAGAGTTTAGAACAAACAGCTCGAGAGGTTTTAACTCTACTGCAAGGGGT
CCAGGATTACAAAGACGATGACGATAAGGCATCCGCTAAGNAAA

SEQ ID NOS:96 and 177

TTAATACGACTCACTATAGGGATTACTATTTACAATTCTTACTTCACAATGCTGGACCCTGTA
AAGGATGTTCTAATTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAG
GATTTGCCTTTTGAGCAAATAGGGAATCTAAGGGAGGAAATTATCAACTGTGCACAAGCGGAT
TACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:97 and 178

TTCTATTTACAATTCTCCTAACACAATGGCCAACCTCCTACTCCTCATGGTACCCATTCTAAT
CGCAATGGCATTCCTAATGCTTACCGAACGAAAAATTCTAGGCTATATACAACCACGCGATTA
CAAAGACGATGACGATAAGGCATCCGCTAAANAAA

SEQ ID NOS:98 and 179

AATTCTCCTAACACANTGCTCCGGCTAAATACTACCGTATGGCCCACCATAATTACCCCCATA
CTCCTTACACTATTCCTCATCACCAACCGACTAATCACCACCCGGGATTACAAAGACGATGAC
GATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:99 and 180

CTATTTACAATTCTCCTAACACAATGACCCTCTACCTAAAACTCACAGCCCTCGCTGTCACTT
TCCTAGGACTTCTAACAGCCCTAGACCTCAACTACCCAACCGATTACAAAGACGATGACGATA
AGGCATCCGCTATNAAAAAA

FIG. 3C-5

SEQ ID NOS:100 and 181

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGCGGGCGTGT
TCTCAGCCGAGCCGTCGCCGTTTCCACAGACCCGTCGCAGCATGGTGTTTGCCAGGCACCTGC
GGGAGGTGGGAGACGAGTTCAGGAGCAGACATCTCAACTCCACGGACGACGCAGACGAGGATT
ACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:101 and 182

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGGCTTAAAAC
TTGCCACAGTTGCTGCCAGTATGGACAGAGTGCCAAAGGTTACTCCCAGCAGTGCCATCAGCA
GCATAGCAAGAGAGAACCACGAACCAGAAAGATTGGGCTTAAATGGAATAGCAGAGACAACAG
ATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:102 and 183

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGATGCGGGATC
TCCCAGGACACTACTATGAAACGCTCAAATTCCTTGTGGGCCATCTCAAGACCATCGCTGACC
ACCGCGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:103 and 184

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTAGGTGTGGATGTGTGG
GGGCCGCATGGAGGACATCCCCTGCTCCAGGGTGGGCCATATCTACAGGAAGTATGTGCCCTA
CAAGGTCCCGGCCGGAGTCAGCCTGGCCCGGAACCTTAAGCGGGTGGCCGATTGGATGGATTA
CAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:104 and 185

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGCGCTTAGTT
GGATCGAAATGGACACCGAGATGGAGATGCTTCTGGCTAGATTTCGCAGAACCCCAGGAGACC
TGCATTTAGACCACTCTGTCCATTTGTGTGCCCACCCCGATTACAAAGACGATGACGATAAGG
CATCCGCTATTTAAAA

SEQ ID NOS:105 and 186

CTATTTACAATTCTCCTAACACAATGACCTCCACCCTACCACACATTCGAAGAACCCGTGATT
ACAAAGACGATGACGATAAGGCATCCGCTATTTAAAAA

SEQ ID NOS:106 and 187

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGAACGGAAATC
TGTTCGCTTCATTCATCGCCGACAGTGATTACAAAGACGATGACGATAAGGCATCCGCTATTT
AAAA

FIG. 3C-6

SEQ ID NOS:107 and 188

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTACTTCGCCCTGGACGACATCG
AGTGGTTTGTGGCCCCGGCTGCAGAAGGCAGCCGAGGCTTTCAAGCAGCTGAACCAGCCCGAT
TACAAAGACCATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:108 and 189

TAATACGACTCCTATAGGGACAATTACTATTTACAATTCTTACTTCAATACAATGCGCACCCT
GCAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCCTACT
CCAATGGCTCAGACTGCACCCGCTGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTA
AAA

SEQ ID NOS:109 and 190

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACGCCAAAGCACAATGGC
TGTTATAATTAACGAATTATCTCAGCGTGACAGCTGTGGTCCTTTGAAAATTAGCTTGAATAA
CAAGATCCTGGTGTATGGTAATTTATTTTCCTCTTTCACCCCCGATTACAAAGACGATGACGA
TAAGGCATCCGCTATTTAAAA

SEQ ID NOS:110 and 191

CAATTCTCCTAACACGATGGGACTGGCTAAAAAAAGTAAAAGGAACCCGGCAAATCTTACCCC
GCCTGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAAA

SEQ ID NOS:111 and 192

NATTTCTATTTACAATTCTCCTAACACAATGAGCTCACAGGCACTTAGAATCCATCAGTGGCT
CCATCTTTTCTCAGACTTCACCTCCACCGATTACAAAGACGATGACGATAAGGCATCCGCTNN
AAAAA

SEQ ID NOS:112 and 193

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGACCAACCCA
TAGGAAAATGGGAAAAGTTGTTCCCGTTACAACTTTACAAAACGTTACAAATGCTCATGTCCC
AGATGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:113 and 194

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTACTTCACAATGGGGGTCTCTG
AGGCCGAGGGAACATTCCCGCTCAGCACTTTCCTTCTTGGGATAGCATCCCGTCTAAGAAGCG
TGGCTGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

FIG. 3C-7

SEQ ID NOS:114 and 195

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGAGGGCGCCCA
GATTCATAAAGCAAATATTGCTAGATCTAAAGAGAGAGATAGACTTCAATGTGAGATTAGTAG
AATACTTCAACCCACTATCAGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:115 and 196

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGATCGTGGCTA
TCATTGCTGGTCGCCTTCGGATGTTGGGTGACCAGTTCAACGGAGAATTGGAAGCTTCTGCCA
AAAACGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:116 and 197

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAACCTGGCTTTGGC
CTATTACAGCAGCCGACAGTATGCTTCAGCACTGAAGCATATCGCTGAGATTATTGAGCGTGG
CATCCGCCAGCACGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:117 and 198

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACGATGGCTGCCATGT
TATTAGACAGAAGAGGAACTGAGTGTGACCTCTGGATAAATGAGATGTCACTATTACATAAGA
TTGTTCAAGATGTATATGGAACTCCTCACCCGCCCCACTCCGATTACAAAGACGATGACGATA
AGGCATCCGCTATTTAAAA

SEQ ID NOS:118 and 199

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGCCTTGGCAAT
ACAAACCGATAGCTGATCTTTACAGAGGGAGAGAGAGCCGTCCCTCTGCCCCCCGGGATTACA
AAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:119 and 200

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGCTGTTCTCAG
TGTTGCTACGTTATTTGGCAGATAACTTTCTGCCAGGAGGATCCGATTACAAAGACGATGACG
ATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:120 and 201

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGATTGGCAGG
TGTTGCTAGGAAAACTACTTTGGAAAATAGATAATCCGGGCATCGATTACAAAGACGATGACG
ATAGGCATCCGCTATTTAAAA

FIG. 3C-8

SEQ ID NOS:121 and 202

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGGTGCTATGG
AGAGAGAATGGGCGATGTTTCTCAGGGCTGCTTCAAGCAGGATTAGGGGTGGCGTGGATTACA
AAGACGATGACGATAAGGCATCCGCTGTTTAAAA

SEQ ID NOS:122 and 203

CTATTTACAATTCTCCTAACACAATGGTGCATAACTTTGGGAGACACTGGGGTCTGCCCTTGA
GTTTTCTTCTCAATTACCCTTTATTCCTCAGTCCGGATTACAAAGACGATGACGATAAGGCAT
CCGCTATTAAAAAA

SEQ ID NOS:123 and 204

TAATACGACTCACTATAGGAAATACTATTTACAATTCTTACTTCACAATGGCTAGCATGGCTC
CAGTGGGGAGAGATGCAGAAACATTGCAAAAGCAAAAGGAAACTATAAAAGCCTTTCTAAAGA
AACTAGAAGCCCTCATGGCAAGCAATGACAATGCCAATAAAACCGATGACAAAGACGATGACG
ATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:124 and 205

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGTGTCGGGAGC
AGGCTGAACTCACTGGGCTCCGCCTGGCAAGCTTGGGGTTGAAGTTTAATAAAATCGTCCATT
CGTCTATGACGCGCGCCATAGAGACCACCGATTACAAAGACGATGACGATAAGGCATCCGCTA
TTTAAAA

SEQ ID NOS:125 and 206

TAATACGACTCACTATAGGGGACAATTACTATTTACAATTCTTACTTCACAATGGGCACTAGA
ATTAGTGATATGCTAAAATTAATTGCAGACACATGGCAGAGAAATTGTTGCCCTGCGGATTAC
AAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:126 and 207

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGGAGCAGGCCA
GTGTTAAGTATGTTATTCTGGATATGTACAGAGCACTCTTGACACTAATGAATACTTCAACAG
CCACAGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:127 and 208

CAATTCTCCTAACACAATGGAAGACCTAGAGAGTGTGTTAATAAGACTGATCAACTGGGCAAA
AGGAAGCCCCATCCCAGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

FIG. 3C-9

SEQ ID NOS:128 and 209

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACAATGAGGCCGGTGT
CCTTTTGCGGGGCTGTTTGGACTCTGAACAGGGCAATAGGAAGGCATTTTGTCCGAGGTAGCA
GGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:129 and 210

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGCACGCGGTGG
TGGCACGTTTGCTTCACATTGGGGCAATCATGTTCCAACGACTAGACTTCATAGAACAATTGT
CTGCACCCCCAGCGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:130 and 211

CTTTTACAATTCTCCTAACACAATGGGCCAAGGTACACTTTGGGGAAGTGGGATGGAAGCATG
GTTGGCAACGGTGTTGAAGGCACTCCCTTGGCACCCCACATACCAGCTGGAGCCGGATTACAA
AGACGATGACGATAAGGCATCCGCTATANAAAA

SEQ ID NOS:131 and 212

TTCTATTTACAATTCTCCTAACACAATGATAGCACAGGCAACGAAAGCAACAATAGACAAATG
GAACTGCATCAAACTTAAAATCTTCTACACCTCAAAGAAAGAAGCCAGCGATTACAAAGACGA
TGACGATAAGGCATCCGCTANT

SEQ ID NOS:132 and 213

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGTGGTGGATG
TGCCAGATTTTATAGTCTGGCTTGAGGAGGCAGTATCTGATTTACATAGAGCCCTCGATTACA
AAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:133 and 214

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGCAGAGGAGAG
GGAATGAATTCCAGCTGAGAGACCTGGCCGATGCATGGGATTTGTCTTCAAGGTCCAGGCAGA
GGGGATGGCAGATGCCAAATTGCAGAAGTCGAAGAGGGCCCGGAGATTACAAAGACGATGACG
ATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:134 and 215

TTTACAATTCTCCTAACACAATGCGGGGCCTGTGGGTGGACAGGGTCCTAGAGGAATGGGGCC
TGGAACCGCGGCAGGATTACAAAGACGATGACGATAAGGCATCCGCTATTAAAAA

FIG. 3C-10

SEQ ID NOS:135 and 216

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTACTCTACAATGTTCGTGAGG
TCTGTTGGCTGGAGGCTGCAGAACATTGGTGATGACATGGACCACGCCATTTGTGGCCATGAT
GTCAGGCTCGGCGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:136 and 217

GCAGTGGACTCAGAAAGCCAACATGTGGCTCCTCCCAGCGCGATTACAAAGACGATGACGATA
AGGCATCCGCTATTTAAAA

SEQ ID NOS:137 and 218

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTTTCTCTACAATGGCGGGTACAC
AGCCACTTATCCTTGCCCAGTTCATGCGTGTTGGAGGTGACGAACTTCTCCACTTCCTGCTCT
GGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAA

SEQ ID NOS:138 and 219

TAATACGACTCACTATAGGGACAATTACTATTTACAATTCTCCTAACACCATGATGGATACCA
TAAAGGGATTTGACCTAATCACTAATTTTCAGGTGGTGGCTGATGCTTTGAACATCTCTTTGC
TGCCCAATCCATTAGCGACAGCGGATTACAAAGACGATGACGATAAGGCATACGCTATTTAAA
A

SEQ ID NOS:139 and 220

TCTATTTACAATTCTCCTAACACAATGGCCACTTGGATGAAAACCCTTCAAGGATTACTGGAT
AGAATTCAGGCTTTCCCCTCCAGCCCCCACGATTACAAAGACGATGACGATAAGGCATCCGCT
ANGAAAAAA

SEQ ID NOS:140 and 221

CTATTTACAATTCTCCTAACACAATGGAAGCTAATAGAAAACAACCGAAACCAAATAATTCAA
GCACTGCTTATTACAATTTTACTGGGGTCTCTATTTTACCCTCCTACAAGCCCCAGATTACAA
AGACGATGACGATAAGGCATCCGCTATAAAAAA

SEQ ID NOS:141 and 222

TTCTATTTACAATTCTCCTAACACAATGGGGCTCACTCACCCACCACATTAACAACATAAAAC
CCTCATCCACACGAGATTACAAAGACGATGACGATAAGGCATCCGCTANAAAAAA

FIG. 3C-11

SEQ ID NOS:142 and 223

TAATACGACTCATATAGGGACAATTACTATTTACAATTCTTACTTCACAATGGTGAGCTGCTG
GCCGATTACTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGCTACGCAATTGAAAA
GCATAGCGGATTACAAAGACGATGACGATAAGGCATCCGCTATTTAAAAAA

FIG. 4

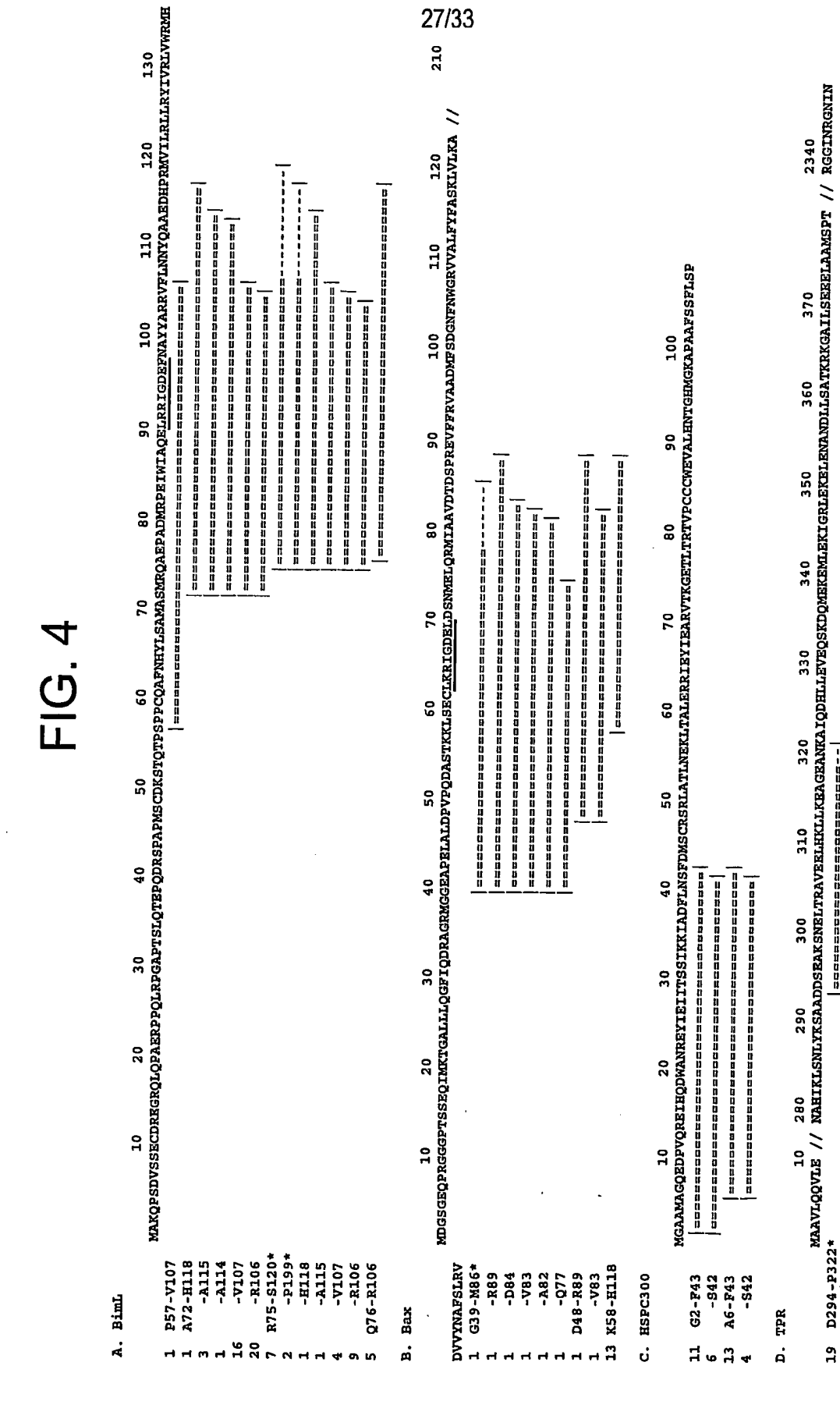


FIG. 5

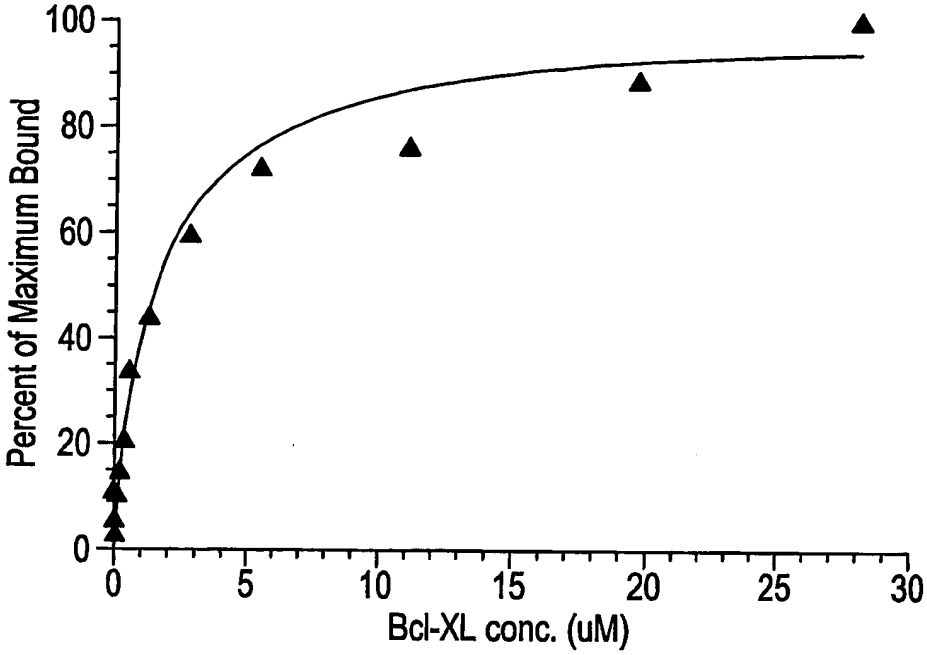

Percent of Maximum Bound
100
80
60
40
20
0
0
5
10
15
20
25
30
Bcl-XL conc. (uM)

FIG. 6

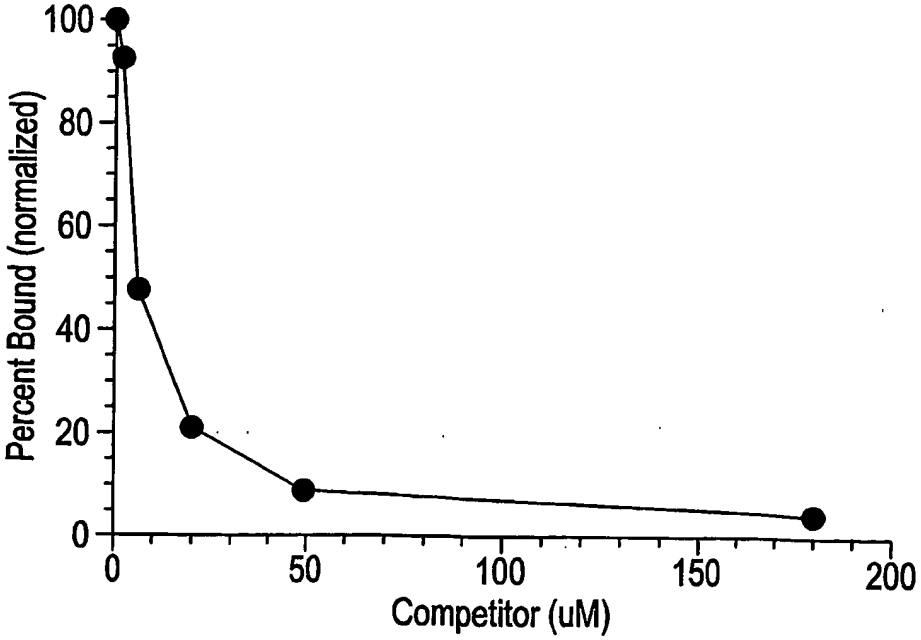

Percent Bound (normalized)
100
80
60
40
20
0
0
50
100
150
200
Competitor (uM)

FIG. 7

Clone	Protein	Comp. Binding	Selected sequence (note that each peptide was preceded by an initiator methione and followed by the sequence DYKDDDDKASA)
f44	Bim	0.84	PEIWIAQELRRIGDEFNAYYARR
T95	Neutrophil Cyto. factor	0.66	QRGMLYYQTEKYDLAIKDL-DEALIQLRGNN
V47	Hypothetical protein	0.36	RWWMCGGRMEDMLCCRVGH
C21	Novel Protein I	0.44	IAQATKATIDKWNCIKLKIFYTSKKEAS
V18	Novel Protein A	0.37	VVDVPDFIVWLEEAVSDLHRAL
X56	Talin	0.21	GGESDTDPHFQDALMQLAKAVASAAAALVLKAKSVAQR
V72	Unknown protein	0.2	AAMLLDRRGTECDLWINEMSLLHKIVQDVYGTPHPPHS
c32	Bak	0.36	GQVGRQLAIIGDDINRRK
Y37	Unknown protein	ND	ALSWIEMDTEMEMLLARFRRTPGDLHLDHSVHLCAHP
Y75	Bcl-2 related protein BPR	ND	TGKEAILRRLVALLEEEAEVINQKLASDPALRSKLVRLSSDSFAHL
V06	Golgi SNAP receptor	0.12	GTRQDRMFETMAIEIEQLLARLTGVNDKMAEYT
C68	HSPC300	0.21	AVQEDPVQREIHQDWANREYIEITTSSIKKIAD
U58	Syntaxin	0.14	ATRQALNEISARHSGIQQLERSIRELHDIFTFL
V50	Tumor protein	0.18	MFSDIYGIREIADGLCLEVEGKMVSRPE
C49	Bax	0.06	KLSECLKRIGDELDSNMELQRMIAAVDTDSPR
U15	Toll-like R3	0.07	FWLEERDFEAGVFELEAIVNSIKRS
Y01	Unknown protein	0.08	LFSVLLRYLADNFLPGGS
W06	Anion Channel	0.32	MKWDTDNTLGTEISWENKLAEGLKLTLDTIFVHHVLHAPH
V68	Novel B	0.9	CREQAELTGLRLASLGLKFNKIVHSSMTRAIET
T25	Unknown D	0.26	GTRISDMLKLIADTWQRNCCPA

FIG. 8A

Number	Protein Name	Accession Number	Clone	Protein Sequence
224	Cdc21	CAA52801	AttB-BrM39	KYQQLFEDIRW
225	SRPg	NP_003124	AttB-Hc-6	IGEEFSRAAEKLYLAV
226	Bmf	NP_277038	AttB-Thy-34	KAEVQIARKLQCIADQFHRLHVL
227	Unknown protein from human chromosome 2 clone RP11-22BG11	AC096559	AttB-Br-M45	MGDVVGFIDELEGAVSDLHRAL
228	Unknown protein from human chromosome 14 clone CTD-301HB and RP11-185P	AC007955	AttB-Thy-38	TLRHWGLQFNTRFGV
229	Sequence/protein not in database	none	AttB-BM-51	SRREEAWDALFRGI
230	Sequence/protein not in database	none	AttB-BM-52	TLREIGDLYTSILGRR

FIG. 8B

SEQ ID NO: 224 Cdc21 (Corresponds to nucleic acid sequence of SEQ ID NO: 231)
KYQQLFEDIRW

SEQ ID NO: 225 SRP9 (Corresponds to nucleic acid sequence of SEQ ID NO: 232)
IGEEFSRAAEKLYLAV

SEQ ID NO: 226 Bmf (Corresponds to nucleic acid sequence of SEQ ID NO: 233)
KAEVQIARKLQCIADQFHRLHVL

SEQ ID NO: 227 Unknown protein from chromosome 2 clone RP11-285P (Corresponds to nucleic acid sequence of SEQ ID NO: 234)
MGDVVGFIDELEGAVSDLHRAL

SEQ ID NO: 228 Unknown protein from chromosome 14 clone CTD-3014H8 and RP11-285P (Corresponds to nucleic acid sequence of SEQ ID NO: 235)
TLRHWGLQFNTRFGV

SEQ ID NO: 229 Sequence/protein not in database (Corresponds to nucleic acid sequence of SEQ ID NO: 236)
SRREEAWDALFRGI

SEQ ID NO: 230 Sequence/protein not in database (Corresponds to nucleic acid sequence of SEQ ID NO: 237)
TLREIGDLYTSILGRR

FIG. 8C

SEQ ID NO: 231
AAATACCAGCAACTTTTTGAAGATTCGGTGG

SEQ ID NO: 232
ATCGGGGAGGAGTTCAGCCGCGCTGCCGAGAAGCTTTACCTCGCTGTT

SEQ ID NO: 233
AAAGCAGAGGTACAGATTGCCCGAAAGCTTCAGTGCATTGCAGACCAGTTCCACCGG
CTTCATGTGCTT

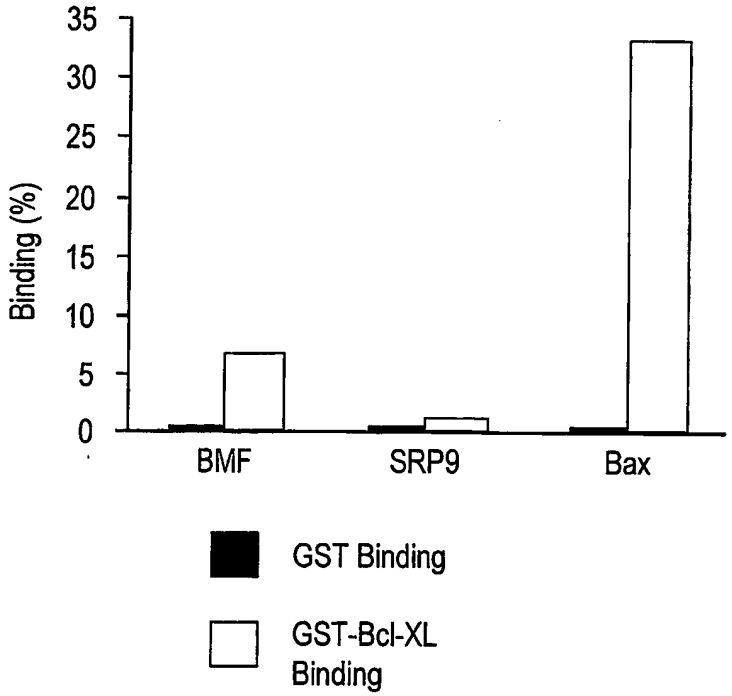
SEQ ID NO: 234
ATGGGAGATGTGGTTGGTTTTATAGACGAACTTGAGGGGGCAGTGTCTGATTTACATA
GGGCGTTG

SEQ ID NO: 235
ACACTCCGACACTGGGGATTACAGTTCAACACAAGATTTGGTGTG

SEQ ID NO: 236
TCGAGAAGGGAAGAGGCATGGGATGCTTTATTTCGTGGGATC

SEQ ID NO: 237
TCGAGAAGGGAAGAGGCATGGGATGCTTTATTTCGTGGGATC

FIG. 9


35
30
25
20
15
10
5
0
Binding (%)
BMF
SRP9
Bax
GST Binding
GST-Bcl-XL Binding

SEQUENCE LISTING

<110> Phylos, Inc.

<120> Polypeptides Interactive with BCL-Xl

<130> 50036/050WO2

<150> US 60/274,526
<151> 2001-03-08

<160> 253

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 35
<212> PRT
<213> Homo sapiens

<400> 1
Ala Ser Met Arg Gln Ala Glu Pro Ala Asp Met Arg Pro Glu Ile Trp
1 5 10 15
Ile Ala Gln Glu Leu Arg Arg Ile Gly Asp Glu Phe Asn Ala Tyr Tyr
20 25 30
Ala Arg Glu
35

<210> 2
<211> 18
<212> PRT
<213> Homo sapiens

<400> 2
Gly Gln Val Gly Arg Gln Leu Ala Ile Ile Gly Asp Asp Ile Asn Arg
1 5 10 15
Arg Lys

<210> 3
<211> 32
<212> PRT
<213> Homo sapiens

<400> 3
Lys Leu Ser Glu Cys Leu Lys Arg Ile Gly Asp Glu Leu Asp Ser Asn
1 5 10 15
Met Glu Leu Gln Arg Met Ile Ala Ala Val Asp Thr Asp Ser Pro Arg
20 25 30

<210> 4
<211> 46
<212> PRT

<213> Homo sapiens

<400> 4
Thr Gly Lys Glu Ala Ile Leu Arg Arg Leu Val Ala Leu Leu Glu Glu
1 5 10 15
Glu Ala Glu Val Ile Asn Gln Lys Leu Ala Ser Asp Pro Ala Leu Arg
20 25 30
Ser Lys Leu Val Arg Leu Ser Ser Asp Ser Phe Ala His Leu
35 40 45

<210> 5
<211> 30
<212> PRT
<213> Homo sapiens

<400> 5
Gln Arg Gly Met Leu Tyr Tyr Gln Thr Glu Lys Tyr Asp Leu Ala Ile
1 5 10 15
Lys Asp Leu Lys Glu Ala Leu Ile Gln Leu Arg Gly Asn Asn
20 25 30

<210> 6
<211> 38
<212> PRT
<213> Homo sapiens

<400> 6
Gly Gly Glu Ser Asp Thr Asp Pro His Phe Gln Asp Ala Leu Met Gln
1 5 10 15
Leu Ala Lys Ala Val Ala Ser Ala Ala Ala Ala Leu Val Leu Lys Ala
20 25 30
Lys Ser Val Ala Gln Arg
35

<210> 7
<211> 35
<212> PRT
<213> Homo sapiens

<400> 7
Gly Thr Arg Gln Asp Arg Met Phe Glu Thr Met Ala Ile Glu Ile Glu
1 5 10 15
Gln Leu Leu Ala Arg Leu Thr Gly Val Asn Asp Lys Met Ala Glu Tyr
20 25 30
Thr Asn Ala
35

<210> 8
<211> 33
<212> PRT
<213> Homo sapiens

<400> 8
Ala Val Gln Glu Asp Pro Val Gln Arg Glu Ile His Gln Asp Trp Ala

```
 1               5                  10                  15
Asn Arg Glu Tyr Ile Glu Ile Ile Thr Ser Ser Ile Lys Lys Ile Ala
            20                  25                  30
Asp
```

<210> 9
<211> 33
<212> PRT
<213> Homo sapiens

<400> 9

```
Ala Thr Arg Gln Ala Leu Asn Glu Ile Ser Ala Arg His Ser Gly Ile
 1               5                  10                  15
Gln Gln Leu Glu Arg Ser Ile Arg Glu Leu His Asp Ile Phe Thr Phe
            20                  25                  30
Leu
```

<210> 10
<211> 28
<212> PRT
<213> Homo sapiens

<400> 10

```
Met Phe Ser Asp Ile Tyr Gly Ile Arg Glu Ile Ala Asp Gly Leu Cys
 1               5                  10                  15
Leu Glu Val Glu Gly Lys Met Val Ser Arg Pro Glu
            20                  25
```

<210> 11
<211> 25
<212> PRT
<213> Homo sapiens

<400> 11

```
Phe Trp Leu Glu Glu Arg Asp Phe Glu Ala Gly Val Phe Glu Leu Glu
 1               5                  10                  15
Ala Ile Val Asn Ser Ile Lys Arg Ser
            20                  25
```

<210> 12
<211> 40
<212> PRT
<213> Homo sapiens

<400> 12

```
Met Lys Trp Asp Thr Asp Asn Thr Leu Gly Thr Glu Ile Ser Trp Glu
 1               5                  10                  15
Asn Lys Leu Ala Glu Gly Leu Lys Leu Thr Leu Asp Thr Ile Phe Val
            20                  25                  30
His His Val Leu His Ala Pro His
        35                  40
```

<210> 13
<211> 31
<212> PRT
<213> Homo sapiens

<400> 13
Arg Gly Ala Val Phe Ser Gln Asp Lys Asp Val Val Gln Glu Ala Thr
1 5 10 15
Lys Val Leu Arg Asn Ala Ala Asp Asn Phe Tyr Ile Asn Asp Arg
20 25 30

<210> 14
<211> 33
<212> PRT
<213> Homo sapiens

<400> 14
Thr Gly Thr Gly Ala Pro Arg Phe Ile Lys Glu Val Gln Glu Leu Asn
1 5 10 15
Ser Ala Leu His Gln Ser Asp Leu Ile Asp Ile Tyr Arg Thr Leu His
20 25 30
Pro

<210> 15
<211> 20
<212> PRT
<213> Homo sapiens

<400> 15
Ser Asn Glu Leu Thr Arg Ala Val Glu Glu Leu His Lys Leu Leu Lys
1 5 10 15
Glu Ala Arg Glu
20

<210> 16
<211> 33
<212> PRT
<213> Homo sapiens

<400> 16
Thr Tyr Trp Asn Leu Leu Pro Pro Lys Arg Pro Ile Lys Glu Val Leu
1 5 10 15
Thr Asp Ile Phe Ala Lys Val Leu Glu Lys Gly Trp Val Asp Ser Arg
20 25 30
Ser

<210> 17
<211> 19
<212> PRT
<213> Homo sapiens

<400> 17

```
Leu Phe Thr Ile Leu Leu Thr Leu Trp Thr Met Arg Cys Ser Ser Thr
 1               5                  10                  15
Pro Ser Gly
```

<210> 18
<211> 28
<212> PRT
<213> Homo sapiens

<400> 18
```
Ala Gly Glu Asp Met Glu Ile Ser Val Lys Glu Leu Arg Thr Ile Leu
 1               5                  10                  15
Asn Arg Ile Ile Ser Lys His Lys Asp Leu Arg Thr
            20                  25
```

<210> 19
<211> 26
<212> PRT
<213> Homo sapiens

<400> 19
```
Gly Leu Arg Glu Glu Ser Glu Glu Tyr Met Ala Ala Ala Asp Glu Tyr
 1               5                  10                  15
Asn Arg Leu Lys Gln Val Lys Gln Pro Ala
            20                  25
```

<210> 20
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> VARIANT
<222> 58, 62, 65
<223> Xaa = Any Amino Acid

<400> 20
```
Lys Gly Ile Ile Ser Arg Leu Met Ser Val Glu Glu Glu Leu Lys Arg
 1               5                  10                  15
Asp His Ala Glu Met Gln Ala Gly Cys Gly Leu Gln Thr Glu Asp His
            20                  25                  30
Leu Met Pro Arg Arg Ser Ala Phe Ala Ser Leu Asp Ala Val Asn Ala
        35                  40                  45
Arg Leu Met Ser Ala Leu Thr Pro Ala Xaa Arg Tyr Val Xaa His Cys
    50                  55                  60
Xaa Pro Leu
65
```

<210> 21
<211> 26
<212> PRT
<213> Homo sapiens

<400> 21
Trp Glu Arg Ile Glu Glu Arg Leu Ala Tyr Ile Ala Asp His Leu Gly
1 5 10 15
Phe Ser Trp Thr Glu Leu Ala Arg Ala Leu
20 25

<210> 22
<211> 27
<212> PRT
<213> Homo sapiens

<400> 22
Ala Arg Gly Asp Phe Ala Gln Ala Ala Gln Gln Leu Trp Leu Ala Leu
1 5 10 15
Arg Ala Leu Gly Arg Pro Leu Pro Thr Ser His
20 25

<210> 23
<211> 30
<212> PRT
<213> Homo sapiens

<400> 23
Gly Ser Ser Lys Asp Leu Ala Lys His Ile Gln Val Val Cys Asp Gly
1 5 10 15
Met Asp Leu Thr Pro Lys Ile His Asp Leu Lys Pro Gln Cys
20 25 30

<210> 24
<211> 33
<212> PRT
<213> Homo sapiens

<400> 24
Gly Phe Leu Ala Ala Glu Gln Asp Ile Arg Glu Glu Ile Arg Lys Val
1 5 10 15
Val Gln Ser Leu Glu Gln Thr Ala Arg Glu Val Leu Thr Leu Leu Gln
20 25 30
Gly

<210> 25
<211> 33
<212> PRT
<213> Homo sapiens

<400> 25
Leu Asp Pro Val Lys Asp Val Leu Ile Leu Ser Ala Leu Arg Arg Met
1 5 10 15
Leu Trp Ala Ala Asp Asp Phe Leu Glu Asp Leu Pro Phe Glu Gln Ile
20 25 30
Gly

<210> 26
<211> 31
<212> PRT
<213> Homo sapiens

<400> 26
Ala Asn Leu Leu Leu Leu Met Val Pro Ile Leu Ile Ala Met Ala Phe
1 5 10 15
Leu Met Leu Thr Glu Arg Lys Ile Leu Gly Tyr Ile Gln Pro Arg
20 25 30

<210> 27
<211> 30
<212> PRT
<213> Homo sapiens

<400> 27
Leu Arg Leu Asn Thr Thr Val Trp Pro Thr Ile Ile Thr Pro Ile Leu
1 5 10 15
Leu Thr Leu Phe Leu Ile Thr Asn Arg Leu Ile Thr Thr Arg
20 25 30

<210> 28
<211> 26
<212> PRT
<213> Homo sapiens

<400> 28
Thr Leu Tyr Leu Lys Leu Thr Ala Leu Ala Val Thr Phe Leu Gly Leu
1 5 10 15
Leu Thr Ala Leu Asp Leu Asn Tyr Pro Thr
20 25

<210> 29
<211> 44
<212> PRT
<213> Homo sapiens

<400> 29
Ala Gly Val Phe Ser Ala Glu Pro Ser Pro Phe Pro Gln Thr Arg Arg
1 5 10 15
Ser Met Val Phe Ala Arg His Leu Arg Glu Val Gly Asp Glu Phe Arg
20 25 30
Ser Arg His Leu Asn Ser Thr Asp Asp Ala Asp Glu
35 40

<210> 30
<211> 45
<212> PRT
<213> Homo sapiens

<400> 30
Gly Leu Lys Leu Ala Thr Val Ala Ala Ser Met Asp Arg Val Pro Lys
1 5 10 15

Val Thr Pro Ser Ser Ala Ile Ser Ser Ile Ala Arg Glu Asn His Glu
20 25 30
Pro Glu Arg Leu Gly Leu Asn Gly Ile Ala Glu Thr Thr
35 40 45

<210> 31
<211> 26
<212> PRT
<213> Homo sapiens

<400> 31
Met Arg Asp Leu Pro Gly His Tyr Tyr Glu Thr Leu Lys Phe Leu Val
1 5 10 15
Gly His Leu Lys Thr Ile Ala Asp His Arg
20 25

<210> 32
<211> 42
<212> PRT
<213> Homo sapiens

<400> 32
Cys Gly Gly Arg Met Glu Asp Ile Pro Cys Ser Arg Val Gly His Ile
1 5 10 15
Tyr Arg Lys Tyr Val Pro Tyr Lys Val Pro Ala Gly Val Ser Leu Ala
20 25 30
Arg Asn Leu Lys Arg Val Ala Asp Trp Met
35 40

<210> 33
<211> 37
<212> PRT
<213> Homo sapiens

<400> 33
Ala Leu Ser Trp Ile Glu Met Asp Thr Glu Met Glu Met Leu Leu Ala
1 5 10 15
Arg Phe Arg Arg Thr Pro Gly Asp Leu His Leu Asp His Ser Val His
20 25 30
Leu Cys Ala His Pro
35

<210> 34
<211> 11
<212> PRT
<213> Homo sapiens

<400> 34
Thr Ser Thr Leu Pro His Ile Arg Arg Thr Arg
1 5 10

<210> 35
<211> 12

<212> PRT
<213> Homo sapiens

<400> 35
Asn Gly Asn Leu Phe Ala Ser Phe Ile Ala Asp Ser
1 5 10

<210> 36
<211> 29
<212> PRT
<213> Homo sapiens

<400> 36
Ile Leu Thr Ser Pro Trp Thr Thr Ser Ser Gly Leu Trp Pro Arg Leu
1 5 10 15
Gln Lys Ala Ala Glu Ala Phe Lys Gln Leu Asn Gln Pro
20 25

<210> 37
<211> 32
<212> PRT
<213> Homo sapiens

<400> 37
Arg Thr Leu Gln Pro Arg Leu Leu Gln Asn Gln Gln Gln His Leu Pro
1 5 10 15
Ala Leu Pro Ile Trp Phe Leu Leu Gln Trp Leu Arg Leu His Pro Leu
20 25 30

<210> 38
<211> 37
<212> PRT
<213> Homo sapiens

<400> 38
Met Ala Val Ile Ile Asn Glu Leu Ser Gln Arg Asp Ser Cys Gly Pro
1 5 10 15
Leu Lys Ile Ser Leu Asn Asn Lys Ile Leu Val Tyr Gly Asn Leu Phe
20 25 30
Ser Ser Phe Thr Pro
35

<210> 39
<211> 16
<212> PRT
<213> Homo sapiens

<400> 39
Gly Leu Ala Lys Lys Ser Lys Arg Asn Pro Ala Asn Leu Thr Pro Pro
1 5 10 15

<210> 40
<211> 20

<212> PRT
<213> Homo sapiens

<400> 40
Ser Ser Gln Ala Leu Arg Ile His Gln Trp Leu His Leu Phe Ser Asp
1 5 10 15
Phe Thr Ser Thr
20

<210> 41
<211> 18
<212> PRT
<213> Homo sapiens

<400> 41
Gly Gln Val Gly Arg Gln Leu Ala Ile Ile Gly Asp Asp Ile Asn Arg
1 5 10 15
Arg Lys

<210> 42
<211> 26
<212> PRT
<213> Homo sapiens

<400> 42
Gly Val Ser Glu Ala Glu Gly Thr Phe Pro Leu Ser Thr Phe Leu Leu
1 5 10 15
Gly Ile Ala Ser Arg Leu Arg Ser Val Ala
20 25

<210> 43
<211> 31
<212> PRT
<213> Homo sapiens

<400> 43
Arg Ala Pro Arg Phe Ile Lys Gln Ile Leu Leu Asp Leu Lys Arg Glu
1 5 10 15
Ile Asp Phe Asn Val Arg Leu Val Glu Tyr Phe Asn Pro Leu Ser
20 25 30

<210> 44
<211> 26
<212> PRT
<213> Homo sapiens

<400> 44
Ile Val Ala Ile Ile Ala Gly Arg Leu Arg Met Leu Gly Asp Gln Phe
1 5 10 15
Asn Gly Glu Leu Glu Ala Ser Ala Lys Asn
20 25

<210> 45
<211> 29
<212> PRT
<213> Homo sapiens

<400> 45
Leu Ala Leu Ala Tyr Tyr Ser Ser Arg Gln Tyr Ala Ser Ala Leu Lys
1 5 10 15
His Ile Ala Glu Ile Ile Glu Arg Gly Ile Arg Gln His
20 25

<210> 46
<211> 38
<212> PRT
<213> Homo sapiens

<400> 46
Ala Ala Met Leu Leu Asp Arg Arg Gly Thr Glu Cys Asp Leu Trp Ile
1 5 10 15
Asn Glu Met Ser Leu Leu His Lys Ile Val Gln Asp Val Tyr Gly Thr
20 25 30
Pro His Pro Pro His Ser
35

<210> 47
<211> 22
<212> PRT
<213> Homo sapiens

<400> 47
Pro Trp Gln Tyr Lys Pro Ile Ala Asp Leu Tyr Arg Gly Arg Glu Ser
1 5 10 15
Arg Pro Ser Ala Pro Arg
20

<210> 48
<211> 18
<212> PRT
<213> Homo sapiens

<400> 48
Leu Phe Ser Val Leu Leu Arg Tyr Leu Ala Asp Asn Phe Leu Pro Gly
1 5 10 15
Gly Ser

<210> 49
<211> 18
<212> PRT
<213> Homo sapiens

<400> 49
Asp Trp Gln Val Leu Leu Gly Lys Leu Leu Trp Lys Ile Asp Asn Pro
1 5 10 15

Gly Ile

<210> 50
<211> 22
<212> PRT
<213> Homo sapiens

<400> 50
Gly Ala Met Glu Arg Glu Trp Ala Met Phe Leu Arg Ala Ala Ser Ser
1 5 10 15
Arg Ile Arg Gly Gly Val
20

<210> 51
<211> 24
<212> PRT
<213> Homo sapiens

<400> 51
Val His Asn Phe Gly Arg His Trp Gly Leu Pro Leu Ser Phe Leu Leu
1 5 10 15
Asn Tyr Pro Leu Phe Leu Ser Pro
20

<210> 52
<211> 40
<212> PRT
<213> Homo sapiens

<400> 52
Ala Ser Met Ala Pro Val Gly Arg Asp Ala Glu Thr Leu Gln Lys Gln
1 5 10 15
Lys Glu Thr Ile Lys Ala Phe Leu Lys Lys Leu Glu Ala Leu Met Ala
20 25 30
Ser Asn Asp Asn Ala Asn Lys Thr
35 40

<210> 53
<211> 33
<212> PRT
<213> Homo sapiens

<400> 53
Cys Arg Glu Gln Ala Glu Leu Thr Gly Leu Arg Leu Ala Ser Leu Gly
1 5 10 15
Leu Lys Phe Asn Lys Ile Val His Ser Ser Met Thr Arg Ala Ile Glu
20 25 30
Thr

<210> 54
<211> 22

<212> PRT
<213> Homo sapiens

<400> 54
Gly Thr Arg Ile Ser Asp Met Leu Lys Leu Ile Ala Asp Thr Trp Gln
1 5 10 15
Arg Asn Cys Cys Pro Ala
20

<210> 55
<211> 26
<212> PRT
<213> Homo sapiens

<400> 55
Glu Gln Ala Ser Val Lys Tyr Val Ile Leu Asp Met Tyr Arg Ala Leu
1 5 10 15
Leu Thr Leu Met Asn Thr Ser Thr Ala Thr
20 25

<210> 56
<211> 20
<212> PRT
<213> Homo sapiens

<400> 56
Glu Asp Leu Glu Ser Val Leu Ile Arg Leu Ile Asn Trp Ala Lys Gly
1 5 10 15
Ser Pro Ile Pro
20

<210> 57
<211> 25
<212> PRT
<213> Homo sapiens

<400> 57
Arg Pro Val Ser Phe Cys Gly Ala Val Trp Thr Leu Asn Arg Ala Ile
1 5 10 15
Gly Arg His Phe Val Arg Gly Ser Arg
20 25

<210> 58
<211> 29
<212> PRT
<213> Homo sapiens

<400> 58
His Ala Val Val Ala Arg Leu Leu His Ile Gly Ala Ile Met Phe Gln
1 5 10 15
Arg Leu Asp Phe Ile Glu Gln Leu Ser Ala Pro Pro Ala
20 25

<210> 59
<211> 31
<212> PRT
<213> Homo sapiens

<400> 59
Gly Gln Gly Thr Leu Trp Gly Ser Gly Met Glu Ala Trp Leu Ala Thr
1 5 10 15
Val Leu Lys Ala Leu Pro Trp His Pro Thr Tyr Gln Leu Glu Pro
20 25 30

<210> 60
<211> 28
<212> PRT
<213> Homo sapiens

<400> 60
Ile Ala Gln Ala Thr Lys Ala Thr Ile Asp Lys Trp Asn Cys Ile Lys
1 5 10 15
Leu Lys Ile Phe Tyr Thr Ser Lys Lys Glu Ala Ser
20 25

<210> 61
<211> 22
<212> PRT
<213> Homo sapiens

<400> 61
Val Val Asp Val Pro Asp Phe Ile Val Trp Leu Glu Glu Ala Val Ser
1 5 10 15
Asp Leu His Arg Ala Leu
20

<210> 62
<211> 39
<212> PRT
<213> Homo sapiens

<400> 62
Gln Arg Arg Gly Asn Glu Phe Gln Leu Arg Asp Leu Ala Asp Ala Trp
1 5 10 15
Asp Leu Ser Ser Arg Ser Arg Gln Arg Gly Trp Gln Met Pro Asn Cys
20 25 30
Arg Ser Arg Arg Gly Pro Gly
35

<210> 63
<211> 18
<212> PRT
<213> Homo sapiens

<400> 63
Arg Gly Leu Trp Val Asp Arg Val Leu Glu Glu Trp Gly Leu Glu Pro
1 5 10 15

Arg Gln

<210> 64
<211> 28
<212> PRT
<213> Homo sapiens

<400> 64
Phe Val Arg Ser Val Gly Trp Arg Leu Gln Asn Ile Gly Asp Asp Met
1 5 10 15
Asp His Ala Ile Cys Gly His Asp Val Arg Leu Gly
20 25

<210> 65
<211> 13
<212> PRT
<213> Homo sapiens

<400> 65
Ser Gly Leu Arg Lys Pro Thr Cys Gly Ser Ser Gln Arg
1 5 10

<210> 66
<211> 25
<212> PRT
<213> Homo sapiens

<400> 66
Ala Gly Thr Gln Pro Leu Ile Leu Ala Gln Phe Met Arg Val Gly Gly
1 5 10 15
Asp Glu Leu Leu His Phe Leu Leu Trp
20 25

<210> 67
<211> 32
<212> PRT
<213> Homo sapiens

<400> 67
Met Asp Thr Ile Lys Gly Phe Asp Leu Ile Thr Asn Phe Gln Val Val
1 5 10 15
Ala Asp Ala Leu Asn Ile Ser Leu Leu Pro Asn Pro Leu Ala Thr Ala
20 25 30

<210> 68
<211> 22
<212> PRT
<213> Homo sapiens

<400> 68
Ala Thr Trp Met Lys Thr Leu Gln Gly Leu Leu Asp Arg Ile Gln Ala
1 5 10 15

Phe Pro Ser Ser Pro His
20

<210> 69
<211> 30
<212> PRT
<213> Homo sapiens

<400> 69
Glu Ala Asn Arg Lys Gln Pro Lys Pro Asn Asn Ser Ser Thr Ala Tyr
1 5 10 15
Tyr Asn Phe Thr Gly Val Ser Ile Leu Pro Ser Tyr Lys Pro
20 25 30

<210> 70
<211> 16
<212> PRT
<213> Homo sapiens

<400> 70
Gly Ser Leu Thr His His Ile Asn Asn Ile Lys Pro Ser Ser Thr Arg
1 5 10 15

<210> 71
<211> 28
<212> PRT
<213> Homo sapiens

<400> 71
Val Ser Cys Trp Pro Ser Tyr Leu Lys Tyr Pro Leu Ser Thr Ala Ser
1 5 10 15
Ala Ser Leu Leu Ala Thr Gln Leu Lys Ser Ile Ala
20 25

<210> 72
<211> 199
<212> DNA
<213> Homo sapiens

<400> 72
taatacgact cactataggg acaattacta tttacaattc ttacttcaca atggcttcca 60
tgaggcaggc tgaacctgca gatatgcgcc cagagatatg gatcgcccaa gagttgcggc 120
gtattggaga cgagtttaac gcctactatg caagggagga ttacaaagac gatgacgata 180
aggcatccgc tatttaaaa 199

<210> 73
<211> 126
<212> DNA
<213> Homo sapiens

<400> 73
tactatttac aattctccta acacaatggg ggcaggtggg gacggcagct cgccatcatc 60
ggggacgaca tcaaccgacg gaaagattac aaagacgatg acgataaggc atccgctatt 120
aaaaaa 126

<210> 74
<211> 160
<212> DNA
<213> Homo sapiens

<400> 74
tttacaattc tcctaacaca atgaagctga gcgagtgtct caagcgcatc ggggacgaac 60
tggacagtaa catggagctg cagaggatga ttgccgccgt ggacacagac tccccccgag 120
attacaaaga cgatgacgat aaggcatccg ctattaaaaa 160

<210> 75
<211> 232
<212> DNA
<213> Homo sapiens

<400> 75
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgacaggga 60
aggaagccat actgcggagg ctggtggccc tgctggagga ggaggcagaa gtcattaacc 120
agaagctggc ctcggacccc gccctgcgca gcaagctggt ccgcctgtcc tccgactctt 180
tcgcccacct ggattacaaa gacgatgacg ataaggcatc cgctatttaa aa 232

<210> 76
<211> 172
<212> DNA
<213> Homo sapiens

<400> 76
gactcactat agggacaatt actatttaca attcttactt ccaacgaggg atgctctact 60
accagacaga gaaatatgat ttggctatca aagaccttaa agaagccttg attcagcttc 120
gagggaacaa tgattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 77
<211> 208
<212> DNA
<213> Homo sapiens

<400> 77
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgggtgggg 60
aaagtgatac tgacccccac ttccaggatg cgctaatgca gctcgccaaa gctgtggcaa 120
gtgctgcagc tgccctggtc ctcaaggcca agagtgtggc ccaacgagat tacaaagacg 180
atgacgatag ggcatccgct atttaaaa 208

<210> 78
<211> 199
<212> DNA
<213> Homo sapiens

<400> 78
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgggaacac 60
gccaagacag aatgtttgag acaatggcga ttgagattga acaacttttg gcaaggctta 120
caggggtaaa tgataaaatg gcagaatata ccaacgctga ttacaaagac gatgacgata 180
aggcatccgc tatttaaaa 199

<210> 79
<211> 181
<212> DNA
<213> Homo sapiens

<400> 79
ctatttacaa ttctcctaac acaatggcgg tacaggagga tccggtgcag cgggagattc 60
accaggactg ggctaaccgg gagtacattg agataatcac cagcagcatc aagaaaatcg 120
cagactttct caactcgttc gattacaaag acgatgacga taaggcatcc gctattaaaa 180
a 181

<210> 80
<211> 208
<212> DNA
<213> Homo sapiens

<400> 80
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atggcgactc 60
gacaggcctt aaatgagatc tcggcccggc acagtgggat ccagcagctt gaacgcagta 120
ttcgtgagct gcacgacata ttcacttttc tggctaccga agtgcgagat tacaaagacg 180
atgacgataa ggcatccgct atttaaaa 208

<210> 81
<211> 178
<212> DNA
<213> Homo sapiens

<400> 81
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgatgttct 60
ccgacatcta cgggatccgg gagatcgcgg acgggttgtg cctggaggtg gaggggaaga 120
tggtcagtag gccagaggat tacaaagacg atgacgataa ggcatccgct atttaaaa 178

<210> 82
<211> 169
<212> DNA
<213> Homo sapiens

<400> 82
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgttttggc 60
tggaagaaag ggactttgag gcgggtgttt ttgaactaga agcaattgtt aacagcatca 120
aaagaagcga ttacaaagac gatgacgata aggcatccgc tatttaaaa 169

<210> 83
<211> 214
<212> DNA
<213> Homo sapiens

<400> 83
taatacgact cactataggg acaattacta tttacaattc ttacttcaat acaatgaaat 60
gggacacaga caatactcta gggacagaaa tctcttggga gaataagttg gctgaagggt 120
tgaaactgac tcttgatacc atatttgtac atcacgtcct gcatgcccca cacgattaca 180
aagacgatga cgataaggca tccgctattt aaaa 214

<210> 84
<211> 187
<212> DNA
<213> Homo sapiens

<400> 84
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgcgggggg 60
cagtgttctc ccaggataag gacgtcgtgc aggaggccac aaaggtgctg aggaatgctg 120
ccgacaactt ctacatcaac gacagggatt acaaagacga tgacgataag gcatccgcta 180
tttaaaa 187

<210> 85
<211> 190
<212> DNA
<213> Homo sapiens

<400> 85
gactcactat agggacaatt actatttaca attctcctaa cacaatgacc ggtacaggag 60
cacccagatt cataaaggaa gtccaggaat tgaactcagc tctacatcaa tcggacctaa 120
tagacatcta cagaactctc caccccgctg attacaaaga cgatgacgat aaggcatccg 180
ctatttaaaa 190

<210> 86
<211> 130
<212> DNA
<213> Homo sapiens

<400> 86
tttacaattc tcctaacaca atgacaaaga gcaatgaact aacccgggca gtagaggaac 60
tacacaaact tttgaaagaa gctagggaag attacaaaga cgatgacgat aaggcatccg 120
ctatttaaaa 130

<210> 87
<211> 199
<212> DNA
<213> Homo sapiens

<400> 87
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgacctact 60
ggaacctgct gccccccaag cggcccatca aagaggtgct gacggacatc tttgccaagg 120
tgctggagaa gggctgggtg gacagccgct ccatccacga ttacaaagac gatgacgata 180
aggcatccgc tatttaaaa 199

<210> 88
<211> 97
<212> DNA
<213> Homo sapiens

<400> 88
ctatttacaa ttctcctaac actatggact atgagatgct cttcaactcc ttcagggatt 60
acaaagacga tgacgataag gcatccgcta ttaaaaa 97

<210> 89
<211> 178
<212> DNA
<213> Homo sapiens

<400> 89
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggccgggg 60
aggacatgga gatcagcgtg aaggagttgc ggacaatcct caataggatc atcagcaaac 120
acaaagacct gcggaccgat tacaaagacg atgacgataa ggcatccgct atttaaaa 178

<210> 90
<211> 172
<212> DNA
<213> Homo sapiens

<400> 90
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgggactaa 60

gagaagaaag tgaagagtac atggctgctg ctgatgaata caatagactg aagcaagtga 120
agcaacctgc agattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 91
<211> 318
<212> DNA
<213> Homo sapiens

<400> 91
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgaagggca 60
tcatcagcag gttgatgtcc gtggaggaag aactgaagag ggaccacgca gagatgcaag 120
cggctgtgga ctccaaacag aagatcattg atgcccagga gaagcgcatt gcctcgttgg 180
atgccgccaa tgcccgcctc atgagtgccc tgacccagct gaaagagagg tacagcatgc 240
aagcccgtaa cggcatctcc cccaccaacc ccgcggatta caaagacgat gacgataagg 300
catccactat ttaaaaaa 318

<210> 92
<211> 172
<212> DNA
<213> Homo sapiens

<400> 92
taatacgact cactataggg acaaatacta tttacaattc tcctaacaca atgtgggaac 60
ggattgagga aaggctggct tatattgctg atcaccttgg cttcagctgg acagaattag 120
caagagcgct ggattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 93
<211> 177
<212> DNA
<213> Homo sapiens

<400> 93
taatacgact cactataggg gacaattact atttacaatt gcttacttca caatggctcg 60
gggagacttt gcccaggctg cccagcagct gtggctggcc ctgcgggcac tgggccggcc 120
cctgcccacc tcccacgatt acaaagacga tgacgataag gcatccgcta tttaaaa 177

<210> 94
<211> 160
<212> DNA
<213> Homo sapiens

<400> 94
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggtggtgg 60
atgtgccaga ttttatagtc tggcttgagg aggcagtatc tgatttacat agggccctcg 120
attacaaaga cgatgacgat aaggcatccg ctatttaaaa 160

<210> 95
<211> 170
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 167
<223> n = A,T,C or G

<400> 95
cttttacaat tctcctaaca caatgggctt tttggctgcc gagcaggaca tccgagagga 60

aatcagaaaa gttgtacaga gtttagaaca aacagctcga gaggttttaa ctctactgca 120
aggggtccag gattacaaag acgatgacga taaggcatcc gctaagnaaa 170

<210> 96
<211> 227
<212> DNA
<213> Homo sapiens

<400> 96
ttaatacgac tcactatagg gattactatt tacaattctt acttcacaat gctggaccct 60
gtaaaggatg ttctaattct ttctgctctg agacgaatgc tatgggctgc agatgacttc 120
ttagaggatt tgccttttga gcaaataggg aatctaaggg aggaaattat caactgtgca 180
caagcggatt acaaagacga tgacgataag gcatccgcta tttaaaa 227

<210> 97
<211> 161
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 158
<223> n = A,T,C or G

<400> 97
ttctatttac aattctccta acacaatggc caacctccta ctcctcatgg tacccattct 60
aatcgcaatg gcattcctaa tgcttaccga acgaaaaatt ctaggctata tacaaccacg 120
cgattacaaa gacgatgacg ataaggcatc cgctaaanaa a 161

<210> 98
<211> 149
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 16
<223> n = A,T,C or G

<400> 98
aattctccta acacantgct ccggctaaat actaccgtat ggcccaccat aattaccccc 60
atactcctta cactattcct catcaccaac cgactaatca ccacccggga ttacaaagac 120
gatgacgata aggcatccgc tatttaaaa 149

<210> 99
<211> 146
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 140
<223> n = A,T,C or G

<400> 99
ctatttacaa ttctcctaac acaatgaccc tctacctaaa actcacagcc ctcgctgtca 60
ctttcctagg acttctaaca gccctagacc tcaactaccc aaccgattac aaagacgatg 120
acgataaggc atccgctatn aaaaaa 146

<210> 100
<211> 226
<212> DNA
<213> Homo sapiens

<400> 100
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atggcgggcg 60
tgttctcagc cgagccgtcg ccgtttccac agacccgtcg cagcatggtg tttgccaggc 120
acctgcggga ggtgggagac gagttcagga gcagacatct caactccacg gacgacgcag 180
acgaggatta caaagacgat gacgataagg catccgctat ttaaaa 226

<210> 101
<211> 229
<212> DNA
<213> Homo sapiens

<400> 101
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgggcttaa 60
aacttgccac agttgctgcc agtatggaca gagtgccaaa ggttactccc agcagtgcca 120
tcagcagcat agcaagagag aaccacgaac cagaaagatt gggcttaaat ggaatagcag 180
agacaacaga ttacaaagac gatgacgata aggcatccgc tatttaaaa 229

<210> 102
<211> 172
<212> DNA
<213> Homo sapiens

<400> 102
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgatgcggg 60
atctcccagg acactactat gaaacgctca aattccttgt gggccatctc aagaccatcg 120
ctgaccaccg cgattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 103
<211> 225
<212> DNA
<213> Homo sapiens

<400> 103
taatacgact cactataggg acaattacta tttacaattc tttctctagg tgtggatgtg 60
tgggggccgc atggaggaca tcccctgctc cagggtgggc catatctaca ggaagtatgt 120
gccctacaag gtcccggccg gagtcagcct ggcccggaac cttaagcggg tggccgattg 180
gatggattac aaagacgatg acgataaggc atccgctatt taaaa 225

<210> 104
<211> 205
<212> DNA
<213> Homo sapiens

<400> 104
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggcgctta 60
gttggatcga aatggacacc gagatggaga tgcttctggc tagatttcgc agaaccccag 120
gagacctgca tttagaccac tctgtccatt tgtgtgccca ccccgattac aaagacgatg 180
acgataaggc atccgctatt taaaa 205

<210> 105
<211> 101
<212> DNA
<213> Homo sapiens

<400> 105
ctatttacaa ttctcctaac acaatgacct ccaccctacc acacattcga agaacccgtg 60
attacaaaga cgatgacgat aaggcatccg ctatttaaaa a 101

<210> 106
<211> 130
<212> DNA
<213> Homo sapiens

<400> 106
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgaacggaa 60
atctgttcgc ttcattcatc gccgacagtg attacaaaga cgatgacgat aaggcatccg 120
ctatttaaaa 130

<210> 107
<211> 164
<212> DNA
<213> Homo sapiens

<400> 107
taatacgact cactataggg acaattacta tttacaattc ttacttcgcc ctggacgaca 60
tcgagtggtt tgtggccccg gctgcagaag gcagccgagg ctttcaagca gctgaaccag 120
cccgattaca aagaccatga cgataaggca tccgctattt aaaa 164

<210> 108
<211> 192
<212> DNA
<213> Homo sapiens

<400> 108
taatacgact cctataggga caattactat ttacaattct tacttcaata caatgcgcac 60
cctgcaaccc aggcttcttc aaaaccaaca acagcacctg ccagccctgc ccatatggtt 120
cctactccaa tggctcagac tgcacccgct ggattacaaa gacgatgacg ataaggcatc 180
cgctatttaa aa 192

<210> 109
<211> 210
<212> DNA
<213> Homo sapiens

<400> 109
taatacgact cactataggg acaattacta tttacaattc tcctaacgcc aaagcacaat 60
ggctgttata attaacgaat tatctcagcg tgacagctgt ggtcctttga aaattagctt 120
gaataacaag atcctggtgt atggtaattt attttcctct ttcacccccg attacaaaga 180
cgatgacgat aaggcatccg ctatttaaaa 210

<210> 110
<211> 109
<212> DNA
<213> Homo sapiens

<400> 110
caattctcct aacacgatgg gactggctaa aaaaagtaaa aggaacccgg caaatcttac 60
cccgcctgat tacaaagacg atgacgataa ggcatccgct atttaaaaa 109

<210> 111
<211> 131
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> 1, 125, 126
<223> n = A,T,C or G

<400> 111
natttctatt tacaattctc ctaacacaat gagctcacag gcacttagaa tccatcagtg 60
gctccatctt ttctcagact tcacctccac cgattacaaa gacgatgacg ataaggcatc 120
cgctnnaaaa a 131

<210> 112
<211> 172
<212> DNA
<213> Homo sapiens

<400> 112
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggaccaac 60
ccataggaaa atgggaaaag ttgttcccgt tacaacttta caaaacgtta caaatgctca 120
tgtcccagat ggattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 113
<211> 172
<212> DNA
<213> Homo sapiens

<400> 113
taatacgact cactataggg acaattacta tttacaattc ttacttcaca atgggggtct 60
ctgaggccga gggaacattc ccgctcagca ctttccttct tgggatagca tcccgtctaa 120
gaagcgtggc tgattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 114
<211> 187
<212> DNA
<213> Homo sapiens

<400> 114
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgagggcgc 60
ccagattcat aaagcaaata ttgctagatc taaagagaga gatagacttc aatgtgagat 120
tagtagaata cttcaaccca ctatcagatt acaaagacga tgacgataag gcatccgcta 180
tttaaaa 187

<210> 115
<211> 172
<212> DNA
<213> Homo sapiens

<400> 115
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgatcgtgg 60
ctatcattgc tggtcgcctt cggatgttgg gtgaccagtt caacggagaa ttggaagctt 120
ctgccaaaaa cgattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 116
<211> 180
<212> DNA
<213> Homo sapiens

<400> 116
taatacgact cactataggg acaattacta tttacaattc tttctctaca acctggcttt 60
ggcctattac agcagccgac agtatgcttc agcactgaag catatcgctg agattattga 120
gcgtggcatc cgccagcacg attacaaaga cgatgacgat aaggcatccg ctatttaaaa 180

<210> 117
<211> 208
<212> DNA
<213> Homo sapiens

<400> 117
taatacgact cactataggg acaattacta tttacaattc tttctctacg atggctgcca 60
tgttattaga cagaagagga actgagtgtg acctctggat aaatgagatg tcactattac 120
ataagattgt tcaagatgta tatggaactc ctcacccgcc ccactccgat tacaaagacg 180
atgacgataa ggcatccgct atttaaaa 208

<210> 118
<211> 160
<212> DNA
<213> Homo sapiens

<400> 118
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgccttggc 60
aatacaaacc gatagctgat ctttacagag ggagagagag ccgtccctct gccccccggg 120
attacaaaga cgatgacgat aaggcatccg ctatttaaaa 160

<210> 119
<211> 148
<212> DNA
<213> Homo sapiens

<400> 119
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgctgttct 60
cagtgttgct acgttatttg gcagataact ttctgccagg aggatccgat tacaaagacg 120
atgacgataa ggcatccgct atttaaaa 148

<210> 120
<211> 147
<212> DNA
<213> Homo sapiens

<400> 120
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atggattggc 60
aggtgttgct aggaaaacta ctttggaaaa tagataatcc gggcatcgat tacaaagacg 120
atgacgatag gcatccgcta tttaaaa 147

<210> 121
<211> 160
<212> DNA
<213> Homo sapiens

<400> 121
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgggtgcta 60
tggagagaga atgggcgatg tttctcaggg ctgcttcaag caggattagg ggtggcgtgg 120
attacaaaga cgatgacgat aaggcatccg ctgtttaaaa 160

<210> 122

<211> 140
<212> DNA
<213> Homo sapiens

<400> 122
ctatttacaa ttctcctaac acaatggtgc ataactttgg gagacactgg ggtctgccct 60
tgagttttct tctcaattac cctttattcc tcagtccgga ttacaaagac gatgacgata 120
aggcatccgc tattaaaaaa 140

<210> 123
<211> 211
<212> DNA
<213> Homo sapiens

<400> 123
taatacgact cactatagga aatactattt acaattctta cttcacaatg gctagcatgg 60
ctccagtggg gagagatgca gaaacattgc aaaagcaaaa ggaaactata aaagcctttc 120
taaagaaact agaagccctc atggcaagca atgacaatgc caataaaacc gatgacaaag 180
acgatgacga taaggcatcc gctatttaaa a 211

<210> 124
<211> 196
<212> DNA
<213> Homo sapiens

<400> 124
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgtgtcggg 60
agcaggctga actcactggg ctccgcctgg caagcttggg gttgaagttt aataaaatcg 120
tccattcgtc tatgacgcgc gccatagaga ccaccgatta caaagacgat gacgataagg 180
catccgctat ttaaaa 196

<210> 125
<211> 161
<212> DNA
<213> Homo sapiens

<400> 125
taatacgact cactataggg gacaattact atttacaatt cttacttcac aatgggcact 60
agaattagtg atatgctaaa attaattgca gacacatggc agagaaattg ttgccctgcg 120
gattacaaag acgatgacga taaggcatcc gctatttaaa a 161

<210> 126
<211> 172
<212> DNA
<213> Homo sapiens

<400> 126
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atggagcagg 60
ccagtgttaa gtatgttatt ctggatatgt acagagcact cttgacacta atgaatactt 120
caacagccac agattacaaa gacgatgacg ataaggcatc cgctatttaa aa 172

<210> 127
<211> 120
<212> DNA
<213> Homo sapiens

<400> 127
caattctcct aacacaatgg aagacctaga gagtgtgtta ataagactga tcaactgggc 60

aaaaggaagc cccatcccag attacaaaga cgatgacgat aaggcatccg ctatttaaaa 120

<210> 128
<211> 169
<212> DNA
<213> Homo sapiens

<400> 128
taatacgact cactataggg acaattacta tttacaattc tcctaacaca atgaggccgg 60
tgtccttttg cggggctgtt tggactctga acagggcaat aggaaggcat tttgtccgag 120
gtagcaggga ttacaaagac gatgacgata aggcatccgc tatttaaaa 169

<210> 129
<211> 181
<212> DNA
<213> Homo sapiens

<400> 129
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgcacgcgg 60
tggtggcacg tttgcttcac attggggcaa tcatgttcca acgactagac ttcatagaac 120
aattgtctgc acccccagcg gattacaaag acgatgacga taaggcatcc gctatttaaa 180
a 181

<210> 130
<211> 159
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 155
<223> n = A,T,C or G

<400> 130
cttttacaat tctcctaaca caatgggcca aggtacactt tggggaagtg ggatggaagc 60
atggttggca acggtgttga aggcactccc ttggcacccc acataccagc tggagccgga 120
ttacaaagac gatgacgata aggcatccgc tatanaaaa 159

<210> 131
<211> 148
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 147
<223> n = A,T,C or G

<400> 131
ttctatttac aattctccta acacaatgat agcacaggca acgaaagcaa caatagacaa 60
atggaactgc atcaaactta aaatcttcta cacctcaaag aaagaagcca gcgattacaa 120
agacgatgac gataaggcat ccgctant 148

<210> 132
<211> 160
<212> DNA
<213> Homo sapiens

<400> 132
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggtggtgg 60
atgtgccaga ttttatagtc tggcttgagg aggcagtatc tgatttacat agagccctcg 120
attacaaaga cgatgacgat aaggcatccg ctatttaaaa 160

<210> 133
<211> 211
<212> DNA
<213> Homo sapiens

<400> 133
taatacgact cactataggg acaattacta tttacaattc tttctctaca atgcagagga 60
gagggaatga attccagctg agagacctgg ccgatgcatg ggatttgtct tcaaggtcca 120
ggcagagggg atggcagatg ccaaattgca gaagtcgaag agggcccgga gattacaaag 180
acgatgacga taaggcatcc gctatttaaa a 211

<210> 134
<211> 118
<212> DNA
<213> Homo sapiens

<400> 134
tttacaattc tcctaacaca atgcggggcc tgtgggtgga cagggtccta gaggaatggg 60
gcctggaacc gcggcaggat tacaaagacg atgacgataa ggcatccgct attaaaaa 118

<210> 135
<211> 179
<212> DNA
<213> Homo sapiens

<400> 135
taatacgact cactataggg acaattacta tttacaattc tttactctac aatgttcgtg 60
aggtctgttg gctggaggct gcagaacatt ggtgatgaca tggaccacgc catttgtggc 120
catgatgtca ggctcggcga ttacaaagac gatgacgata aggcatccgc tatttaaaa 179

<210> 136
<211> 82
<212> DNA
<213> Homo sapiens

<400> 136
gcagtggact cagaaagcca acatgtggct cctcccagcg cgattacaaa gacgatgacg 60
ataaggcatc cgctatttaa aa 82

<210> 137
<211> 169
<212> DNA
<213> Homo sapiens

<400> 137
taatacgact cactataggg acaattacta tttacaattc tttctctaca atggcgggta 60
cacagccact tatccttgcc cagttcatgc gtgttggagg tgacgaactt ctccacttcc 120
tgctctggga ttacaaagac gatgacgata aggcatccgc tatttaaaa 169

<210> 138
<211> 190
<212> DNA
<213> Homo sapiens

<400> 138
taatacgact cactataggg acaattacta tttacaattc tcctaacacc atgatggata 60
ccataaaggg atttgaccta atcactaatt ttcaggtggt ggctgatgct ttgaacatct 120
ctttgctgcc caatccatta gcgacagcgg attacaaaga cgatgacgat aaggcatacg 180
ctatttaaaa 190

<210> 139
<211> 135
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 128
<223> n = A,T,C or G

<400> 139
tctatttaca attctcctaa cacaatggcc acttggatga aaacccttca aggattactg 60
gatagaattc aggctttccc ctccagcccc cacgattaca aagacgatga cgataaggca 120
tccgctanga aaaaa 135

<210> 140
<211> 159
<212> DNA
<213> Homo sapiens

<400> 140
ctatttacaa ttctcctaac acaatggaag ctaatagaaa acaaccgaaa ccaaataatt 60
caagcactgc ttattacaat tttactgggg tctctatttt accctcctac aagccccaga 120
ttacaaagac gatgacgata aggcatccgc tataaaaaa 159

<210> 141
<211> 118
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 112
<223> n = A,T,C or G

<400> 141
ttctatttac aattctccta acacaatggg gctcactcac ccaccacatt aacaacataa 60
aaccctcatc cacacgagat tacaaagacg atgacgataa ggcatccgct anaaaaaa 118

<210> 142
<211> 177
<212> DNA
<213> Homo sapiens

<400> 142
taatacgact catataggga caattactat ttacaattct tacttcacaa tggtgagctg 60
ctggccgatt actaaaatac cctttgtcta cagcctccgc ttctctcctg gctacgcaat 120
tgaaaagcat agcggattac aaagacgatg acgataaggc atccgctatt taaaaaa 177

<210> 143
<211> 71
<212> DNA

<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 143
taatacgact cactataggg acaattacta tttacaatth hhhhhhhaca atggctgaag 60
aacagaaact g 71

<210> 144
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 144
taatacgact cactataggg acaattacta tttacaatt 39

<210> 145
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 25, 26, 27, 28, 29, 30, 31, 32, 33
<223> n = A,T,C or G

<223> Oligonucleotide Primer

<400> 145
ggaacttgct tcgtctttgc aatcnnnnnn nnn 33

<210> 146
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 25, 26, 27, 28, 29, 30, 31, 32, 33
<223> n = A,T,C or G

<223> Oligonucleotide Primer

<400> 146
ggatgatgct tcgtctttgt aatcnnnnnn nnn 33

<210> 147
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 36, 37, 38, 39, 40, 41, 42, 43, 44, 45

<223> n = A,T,C or G

<223> Oligonucleotide Primer

<400> 147
ggacaattac tatttacaat thhhhhhhha caatgnnnnn nnnnn 45

<210> 148
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 148
taatacgact cactataggg acaattacta tttacaatt 39

<210> 149
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 149
ttttaaatag cgcatgcctt atcgtcatcg tctttgtaat c 41

<210> 150
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 150
agtatcgaat tcatgtctca gagcaaccgg 30

<210> 151
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide Primer

<400> 151
tacagtctcg agctagttga agcgttcctg gccct 35

<210> 152
<211> 28
<212> PRT
<213> Homo sapiens

<400> 152
Met Gly Gln Val Gly Arg Gln Leu Ala Ile Ile Gly Asp Asp Ile Asn

1 5 10 15
Arg Asp Tyr Lys Asp Asp Asp Asp Lys Ala Ser Ala
20 25

<210> 153
<211> 105
<212> DNA
<213> Homo sapiens

<400> 153
gcttccatga ggcaggctga acctgcagat atgcgcccag agatatggat cgcccaagag 60
ttgcggcgta ttggagacga gtttaacgcc tactatgcaa gggag 105

<210> 154
<211> 56
<212> DNA
<213> Homo sapiens

<400> 154
ggggcaggtg gggacggcag ctcgccatca tcggggacga catcaaccga cggaaa 56

<210> 155
<211> 96
<212> DNA
<213> Homo sapiens

<400> 155
aagctgagcg agtgtctcaa gcgcatcggg gacgaactgg acagtaacat ggagctgcag 60
aggatgattg ccgccgtgga cacagactcc ccccga 96

<210> 156
<211> 138
<212> DNA
<213> Homo sapiens

<400> 156
acagggaagg aagccatact gcggaggctg gtggccctgc tggaggagga ggcagaagtc 60
attaaccaga agctggcctc ggaccccgcc ctgcgcagca agctggtccg cctgtcctcc 120
gactctttcg cccacctg 138

<210> 157
<211> 78
<212> DNA
<213> Homo sapiens

<400> 157
ctctactacc agacagagaa atatgatttg gctatcaaag accttaaaga agccttgatt 60
cagcttcgag ggaacaat 78

<210> 158
<211> 114
<212> DNA
<213> Homo sapiens

<400> 158
ggtggggaaa gtgatactga cccccacttc caggatgcgc taatgcagct cgccaaagct 60
gtggcaagtg ctgcagctgc cctggtcctc aaggccaaga gtgtggccca acga 114

<210> 159
<211> 105
<212> DNA
<213> Homo sapiens

<400> 159
ggaacacgcc aagacagaat gtttgagaca atggcgattg agattgaaca acttttggca 60
aggcttacag gggtaaatga taaaatggca gaatatacca acgct 105

<210> 160
<211> 114
<212> DNA
<213> Homo sapiens

<400> 160
gcggtacagg aggatccggt gcagcgggag attcaccagg actgggctaa ccgggagtac 60
attgagataa tcaccagcag catcaagaaa atcgcagact ttctcaactc gttc 114

<210> 161
<211> 114
<212> DNA
<213> Homo sapiens

<400> 161
gcgactcgac aggccttaaa tgagatctcg gcccggcaca gtgggatcca gcagcttgaa 60
cgcagtattc gtgagctgca cgacatattc acttttctgg ctaccgaagt gcga 114

<210> 162
<211> 84
<212> DNA
<213> Homo sapiens

<400> 162
atgttctccg acatctacgg gatccgggag atcgcggacg ggttgtgcct ggaggtggag 60
gggaagatgg tcagtaggcc agag 84

<210> 163
<211> 75
<212> DNA
<213> Homo sapiens

<400> 163
ttttggctgg aagaaaggga ctttgaggcg ggtgtttttg aactagaagc aattgttaac 60
agcatcaaaa gaagc 75

<210> 164
<211> 117
<212> DNA
<213> Homo sapiens

<400> 164
aaatgggaca cagacaatac tctagggaca gaaatctctt gggagaataa gttggctgaa 60
gggttgaaac tgactcttga taccatattt gtacatcacg tcctgcatgc cccacac 117

<210> 165
<211> 93
<212> DNA
<213> Homo sapiens

<400> 165
cgggggggcag tgttctccca ggataaggac gtcgtgcagg aggccacaaa ggtgctgagg 60
aatgctgccg acaacttcta catcaacgac agg 93

<210> 166
<211> 102
<212> DNA
<213> Homo sapiens

<400> 166
accggtacag gagcacccag attcataaag gaagtccagg aattgaactc agctctacat 60
caatcggacc taatagacat ctacagaact ctccaccccg ct 102

<210> 167
<211> 66
<212> DNA
<213> Homo sapiens

<400> 167
acaaagagca atgaactaac ccgggcagta gaggaactac acaaactttt gaaagaagct 60
agggaa 66

<210> 168
<211> 105
<212> DNA
<213> Homo sapiens

<400> 168
acctactgga acctgctgcc ccccaagcgg cccatcaaag aggtgctgac ggacatcttt 60
gccaaggtgc tggagaaggg ctgggtggac agccgctcca tccac 105

<210> 169
<211> 30
<212> DNA
<213> Homo sapiens

<400> 169
gactatgaga tgctcttcaa ctccttcagg 30

<210> 170
<211> 84
<212> DNA
<213> Homo sapiens

<400> 170
gccggggagg acatggagat cagcgtgaag gagttgcgga caatcctcaa taggatcatc 60
agcaaacaca aagacctgcg gacc 84

<210> 171
<211> 78
<212> DNA
<213> Homo sapiens

<400> 171
ggactaagag aagaaagtga agagtacatg gctgctgctg atgaatacaa tagactgaag 60
caagtgaagc aacctgca 78

<210> 172

<211> 222
<212> DNA
<213> Homo sapiens

<400> 172
aagggcatca tcagcaggtt gatgtccgtg gaggaagaac tgaagaggga ccacgcagag 60
atgcaagcgg ctgtggactc caaacagaag atcattgatg cccaggagaa gcgcattgcc 120
tcgttggatg ccgccaatgc ccgcctcatg agtgccctga cccagctgaa agagaggtac 180
agcatgcaag cccgtaacgg catctccccc accaaccccg cg 222

<210> 173
<211> 78
<212> DNA
<213> Homo sapiens

<400> 173
tgggaacgga ttgaggaaag gctggcttat attgctgatc accttggctt cagctggaca 60
gaattagcaa gagcgctg 78

<210> 174
<211> 81
<212> DNA
<213> Homo sapiens

<400> 174
gctcggggag actttgccca ggctgcccag cagctgtggc tggccctgcg ggcactgggc 60
cggcccctgc ccacctccca c 81

<210> 175
<211> 66
<212> DNA
<213> Homo sapiens

<400> 175
gtggtggatg tgccagattt tatagtctgg cttgaggagg cagtatctga tttacatagg 60
gccctc 66

<210> 176
<211> 105
<212> DNA
<213> Homo sapiens

<400> 176
ggctttttgg ctgccgagca ggacatccga gaggaaatca gaaaagttgt acagagttta 60
gaacaaacag ctcgagaggt tttaactcta ctgcaagggg tccag 105

<210> 177
<211> 135
<212> DNA
<213> Homo sapiens

<400> 177
ctggaccctg taaaggatgt tctaattctt tctgctctga gacgaatgct atgggctgca 60
gatgacttct tagaggattt gccttttgag caaataggga atctaaggga ggaaattatc 120
aactgtgcac aagcg 135

<210> 178
<211> 93

<212> DNA
<213> Homo sapiens

<400> 178
gccaacctcc tactcctcat ggtacccatt ctaatcgcaa tggcattcct aatgcttacc 60
gaacgaaaaa ttctaggcta tatacaacca cgc 93

<210> 179
<211> 90
<212> DNA
<213> Homo sapiens

<400> 179
ctccggctaa atactaccgt atggcccacc ataattaccc ccatactcct tacactattc 60
ctcatcacca accgactaat caccacccgg 90

<210> 180
<211> 78
<212> DNA
<213> Homo sapiens

<400> 180
accctctacc taaaactcac agccctcgct gtcactttcc taggacttct aacagcccta 60
gacctcaact acccaacc 78

<210> 181
<211> 132
<212> DNA
<213> Homo sapiens

<400> 181
gcgggcgtgt tctcagccga gccgtcgccg tttccacaga cccgtcgcag catggtgttt 60
gccaggcacc tgcgggaggt gggagacgag ttcaggagca gacatctcaa ctccacggac 120
gacgcagacg ag 132

<210> 182
<211> 135
<212> DNA
<213> Homo sapiens

<400> 182
ggcttaaaac ttgccacagt tgctgccagt atggacagag tgccaaaggt tactcccagc 60
agtgccatca gcagcatagc aagagagaac cacgaaccag aaagattggg cttaaatgga 120
atagcagaga caaca 135

<210> 183
<211> 78
<212> DNA
<213> Homo sapiens

<400> 183
atgcgggatc tcccaggaca ctactatgaa acgctcaaat tccttgtggg ccatctcaag 60
accatcgctg accaccgc 78

<210> 184
<211> 126
<212> DNA
<213> Homo sapiens

<400> 184
tgtgggggcc gcatggagga catcccctgc tccagggtgg gccatatcta caggaagtat 60
gtgccctaca aggtcccggc cggagtcagc ctggcccgga accttaagcg ggtggccgat 120
tggatg 126

<210> 185
<211> 111
<212> DNA
<213> Homo sapiens

<400> 185
gcgcttagtt ggatcgaaat ggacaccgag atggagatgc ttctggctag atttcgcaga 60
accccaggag acctgcattt agaccactct gtccatttgt gtgcccaccc c 111

<210> 186
<211> 33
<212> DNA
<213> Homo sapiens

<400> 186
acctccaccc taccacacat tcgaagaacc cgt 33

<210> 187
<211> 36
<212> DNA
<213> Homo sapiens

<400> 187
aacggaaatc tgttcgcttc attcatcgcc gacagt 36

<210> 188
<211> 70
<212> DNA
<213> Homo sapiens

<400> 188
gacgacatcg agtggtttgt ggccccggct gcagaaggca gccgaggctt tcaagcagct 60
gaaccagccc 70

<210> 189
<211> 96
<212> DNA
<213> Homo sapiens

<400> 189
cgcaccctgc aacccaggct tcttcaaaac caacaacagc acctgccagc cctgcccata 60
tggttcctac tccaatggct cagactgcac ccgctg 96

<210> 190
<211> 108
<212> DNA
<213> Homo sapiens

<400> 190
gctgttataa ttaacgaatt atctcagcgt gacagctgtg gtcctttgaa aattagcttg 60
aataacaaga tcctggtgta tggtaattta ttttcctctt tcaccccc 108

<210> 191

<211> 48
<212> DNA
<213> Homo sapiens

<400> 191
ggactggcta aaaaaagtaa aaggaacccg gcaaatctta ccccgcct 48

<210> 192
<211> 60
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 1, 125, 126
<223> n = A,T,C or G

<400> 192
agctcacagg cacttagaat ccatcagtgg ctccatcttt tctcagactt cacctccacc 60

<210> 193
<211> 78
<212> DNA
<213> Homo sapiens

<400> 193
gaccaaccca taggaaaatg ggaaaagttg ttcccgttac aactttacaa aacgttacaa 60
atgctcatgt cccagatg 78

<210> 194
<211> 78
<212> DNA
<213> Homo sapiens

<400> 194
ggggtctctg aggccgaggg aacattcccg ctcagcactt tccttcttgg gatagcatcc 60
cgtctaagaa gcgtggct 78

<210> 195
<211> 93
<212> DNA
<213> Homo sapiens

<400> 195
agggcgccca gattcataaa gcaaatattg ctagatctaa agagagagat agacttcaat 60
gtgagattag tagaatactt caacccacta tca 93

<210> 196
<211> 78
<212> DNA
<213> Homo sapiens

<400> 196
atcgtggcta tcattgctgg tcgccttcgg atgttgggtg accagttcaa cggagaattg 60
gaagcttctg ccaaaaac 78

<210> 197

<211> 84
<212> DNA
<213> Homo sapiens

<400> 197
gctttggcct attacagcag ccgacagtat gcttcagcac tgaagcatat cgctgagatt 60
attgagcgtg gcatccgcca gcac 84

<210> 198
<211> 114
<212> DNA
<213> Homo sapiens

<400> 198
gctgccatgt tattagacag aagaggaact gagtgtgacc tctggataaa tgagatgtca 60
ctattacata agattgttca agatgtatat ggaactcctc acccgcccca ctcc 114

<210> 199
<211> 66
<212> DNA
<213> Homo sapiens

<400> 199
ccttggcaat acaaaccgat agctgatctt tacagaggga gagagagccg tccctctgcc 60
ccccgg 66

<210> 200
<211> 54
<212> DNA
<213> Homo sapiens

<400> 200
ctgttctcag tgttgctacg ttatttggca gataactttc tgccaggagg atcc 54

<210> 201
<211> 54
<212> DNA
<213> Homo sapiens

<400> 201
gattggcagg tgttgctagg aaaactactt tggaaaatag ataatccggg catc 54

<210> 202
<211> 66
<212> DNA
<213> Homo sapiens

<400> 202
ggtgctatgg agagagaatg ggcgatgttt ctcagggctg cttcaagcag gattaggggt 60
ggcgtg 66

<210> 203
<211> 72
<212> DNA
<213> Homo sapiens

<400> 203
gtgcataact ttgggagaca ctggggtctg cccttgagtt ttcttctcaa ttacccttta 60

ttcctcagtc cg 72

<210> 204
<211> 120
<212> DNA
<213> Homo sapiens

<400> 204
gctagcatgg ctccagtggg gagagatgca gaaacattgc aaaagcaaaa ggaaactata 60
aaagcctttc taaagaaact agaagccctc atggcaagca atgacaatgc caataaaacc 120

<210> 205
<211> 102
<212> DNA
<213> Homo sapiens

<400> 205
tgtcgggagc aggctgaact cactgggctc cgcctggcaa gcttggggtt gaagtttaat 60
aaaatcgtcc attcgtctat gacgcgcgcc atagagacca cc 102

<210> 206
<211> 66
<212> DNA
<213> Homo sapiens

<400> 206
ggcactagaa ttagtgatat gctaaaatta attgcagaca catggcagag aaattgttgc 60
cctgcg 66

<210> 207
<211> 78
<212> DNA
<213> Homo sapiens

<400> 207
gagcaggcca gtgttaagta tgttattctg gatatgtaca gagcactctt gacactaatg 60
aatacttcaa cagccaca 78

<210> 208
<211> 60
<212> DNA
<213> Homo sapiens

<400> 208
gaagacctag agagtgtgtt aataagactg atcaactggg caaaaggaag ccccatccca 60

<210> 209
<211> 75
<212> DNA
<213> Homo sapiens

<400> 209
aggccggtgt ccttttgcgg ggctgtttgg actctgaaca gggcaatagg aaggcatttt 60
gtccgaggta gcagg 75

<210> 210

<211> 87
<212> DNA
<213> Homo sapiens

<400> 210
cacgcggtgg tggcacgttt gcttcacatt ggggcaatca tgttccaacg actagacttc 60
atagaacaat tgtctgcacc cccagcg 87

<210> 211
<211> 93
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 155
<223> n = A,T,C or G

<400> 211
ggccaaggta cactttgggg aagtgggatg gaagcatggt tggcaacggt gttgaaggca 60
ctcccttggc accccacata ccagctggag ccg 93

<210> 212
<211> 84
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 147
<223> n = A,T,C or G

<400> 212
atagcacagg caacgaaagc aacaatagac aaatggaact gcatcaaact taaaatcttc 60
tacacctcaa agaaagaagc cagc 84

<210> 213
<211> 66
<212> DNA
<213> Homo sapiens

<400> 213
gtggtggatg tgccagattt tatagtctgg cttgaggagg cagtatctga tttacataga 60
gccctc 66

<210> 214
<211> 117
<212> DNA
<213> Homo sapiens

<400> 214
cagaggagag ggaatgaatt ccagctgaga gacctggccg atgcatggga tttgtcttca 60
aggtccaggc agaggggatg gcagatgcca aattgcagaa gtcgaagagg gcccgga 117

<210> 215
<211> 54
<212> DNA
<213> Homo sapiens

<400> 215
cggggcctgt gggtggacag ggtcctagag gaatggggcc tggaaccgcg gcag 54

<210> 216
<211> 84
<212> DNA
<213> Homo sapiens

<400> 216
ttcgtgaggt ctgttggctg gaggctgcag aacattggtg atgacatgga ccacgccatt 60
tgtggccatg atgtcaggct cggc 84

<210> 217
<211> 39
<212> DNA
<213> Homo sapiens

<400> 217
agtggactca gaaagccaac atgtggctcc tcccagcgc 39

<210> 218
<211> 75
<212> DNA
<213> Homo sapiens

<400> 218
gcgggtacac agccacttat ccttgcccag ttcatgcgtg ttggaggtga cgaacttctc 60
cacttcctgc tctgg 75

<210> 219
<211> 96
<212> DNA
<213> Homo sapiens

<400> 219
atggatacca taaagggatt tgacctaatc actaattttc aggtggtggc tgatgctttg 60
aacatctctt tgctgcccaa tccattagcg acagcg 96

<210> 220
<211> 66
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 128
<223> n = A,T,C or G

<400> 220
gccacttgga tgaaaaccct tcaaggatta ctggatagaa ttcaggcttt cccctccagc 60
ccccac 66

<210> 221
<211> 92
<212> DNA
<213> Homo sapiens

<400> 221

gaagctaata gaaaacaacc gaaaccaaat aattcaagca ctgcttatta caattttact 60
ggggtctcta ttttaccctc ctacaagccc ca 92

<210> 222
<211> 49
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 112
<223> n = A,T,C or G

<400> 222
gggctcactc acccaccaca ttaacaacat aaaaccctca tccacacga 49

<210> 223
<211> 82
<212> DNA
<213> Homo sapiens

<400> 223
gtgagctgct ggccgattac taaaataccc tttgtctaca gcctccgctt ctctcctggc 60
tacgcaattg aaaagcatag cg 82

<210> 224
<211> 11
<212> PRT
<213> Homo sapiens

<400> 224
Lys Tyr Gln Gln Leu Phe Glu Asp Ile Arg Trp
1 5 10

<210> 225
<211> 16
<212> PRT
<213> Homo sapiens

<400> 225
Ile Gly Glu Glu Phe Ser Arg Ala Ala Glu Lys Leu Tyr Leu Ala Val
1 5 10 15

<210> 226
<211> 23
<212> PRT
<213> Homo sapiens

<400> 226
Lys Ala Glu Val Gln Ile Ala Arg Lys Leu Gln Cys Ile Ala Asp Gln
1 5 10 15
Phe His Arg Leu His Val Leu
20

<210> 227

<211> 22
<212> PRT
<213> Homo sapiens

<400> 227
Met Gly Asp Val Val Gly Phe Ile Asp Glu Leu Glu Gly Ala Val Ser
1 5 10 15
Asp Leu His Arg Ala Leu
20

<210> 228
<211> 15
<212> PRT
<213> Homo sapiens

<400> 228
Thr Leu Arg His Trp Gly Leu Gln Phe Asn Thr Arg Phe Gly Val
1 5 10 15

<210> 229
<211> 14
<212> PRT
<213> Homo sapiens

<400> 229
Ser Arg Arg Glu Glu Ala Trp Asp Ala Leu Phe Arg Gly Ile
1 5 10

<210> 230
<211> 17
<212> PRT
<213> Homo sapiens

<400> 230
Thr Leu Arg Glu Ile Gly Asp Leu Tyr Leu Thr Ser Ile Leu Gly Arg
1 5 10 15
Arg

<210> 231
<211> 33
<212> DNA
<213> Homo sapiens

<400> 231
aaataccagc aacttttga agatattcgg tgg 33

<210> 232
<211> 48
<212> DNA
<213> Homo sapiens

<400> 232
atcggggagg agttcagccg cgctgccgag aagctttacc tcgctgtt 48

<210> 233
<211> 69
<212> DNA
<213> Homo sapiens

<400> 233
aaagcagagg tacagattgc ccgaaagctt cagtgcattg cagaccagtt ccaccggctt 60
catgtgctt 69

<210> 234
<211> 66
<212> DNA
<213> Homo sapiens

<400> 234
atgggagatg tggttggttt tatagacgaa cttgaggggg cagtgtctga tttacatagg 60
gcgttg 66

<210> 235
<211> 45
<212> DNA
<213> Homo sapiens

<400> 235
acactccgac actggggatt acagttcaac acaagatttg gtgtg 45

<210> 236
<211> 42
<212> DNA
<213> Homo sapiens

<400> 236
tcgagaaggg aagaggcatg ggatgcttta tttcgtggga tc 42

<210> 237
<211> 42
<212> DNA
<213> Homo sapiens

<400> 237
tcgagaaggg aagaggcatg ggatgcttta tttcgtggga tc 42

<210> 238
<211> 18
<212> PRT
<213> Homo sapiens

<400> 238
Met Pro Val Val His Leu Thr Leu Thr Thr Ala Gly Asp Asp Phe Ser
1 5 10 15
Arg Arg

<210> 239
<211> 25
<212> PRT
<213> Homo sapiens

<400> 239
Met Pro Gln Asp Ala Ser Thr Lys Lys Leu Ser Glu Cys Leu Lys Arg
1 5 10 15
Ile Gly Asp Glu Leu Asp Ser Asn Gly
20 25

<210> 240
<211> 17
<212> PRT
<213> Homo sapiens

<400> 240
Met Gly Gln Val Gly Arg Gln Leu Ala Ile Ile Gly Asp Asp Ile Asn
1 5 10 15
Arg

<210> 241
<211> 138
<212> PRT
<213> Homo sapiens

<400> 241
Met Ala Lys Gln Pro Ser Asp Val Ser Ser Glu Cys Asp Arg Glu Gly
1 5 10 15
Arg Gln Leu Gln Pro Ala Glu Arg Pro Pro Gln Leu Arg Pro Gly Ala
20 25 30
Pro Thr Ser Leu Gln Thr Glu Pro Gln Asp Arg Ser Pro Ala Pro Met
35 40 45
Ser Cys Asp Lys Ser Thr Gln Thr Pro Ser Pro Pro Cys Gln Ala Phe
50 55 60
Asn His Tyr Leu Ser Ala Met Ala Ser Met Arg Gln Ala Glu Pro Ala
65 70 75 80
Asp Met Arg Pro Glu Ile Trp Ile Ala Gln Glu Leu Arg Arg Ile Gly
85 90 95
Asp Glu Phe Asn Ala Tyr Tyr Ala Arg Arg Val Phe Leu Asn Asn Tyr
100 105 110
Gln Ala Ala Glu Asp His Pro Arg Met Val Ile Leu Arg Leu Leu Arg
115 120 125
Tyr Ile Val Arg Leu Val Trp Arg Met His
130 135

<210> 242
<211> 135
<212> PRT
<213> Homo sapiens

<400> 242
Met Asp Gly Ser Gly Glu Gln Pro Arg Gly Gly Gly Pro Thr Ser Ser
1 5 10 15
Glu Gln Ile Met Lys Thr Gly Ala Leu Leu Leu Gln Gly Phe Ile Gln
20 25 30
Asp Arg Ala Gly Arg Met Gly Gly Glu Ala Pro Glu Leu Ala Leu Asp
35 40 45
Pro Val Pro Gln Asp Ala Ser Thr Lys Lys Leu Ser Glu Cys Leu Lys

```
    50                  55                  60
Arg Ile Gly Asp Glu Leu Asp Ser Asn Met Glu Leu Gln Arg Met Ile
65                  70                  75                  80
Ala Ala Val Asp Thr Asp Ser Pro Arg Glu Val Phe Phe Arg Val Ala
                85                  90                  95
Ala Asp Met Phe Ser Asp Gly Asn Phe Asn Trp Gly Arg Val Val Ala
            100                 105                 110
Leu Phe Tyr Phe Ala Ser Lys Leu Val Leu Lys Ala Asp Val Val Tyr
        115                 120                 125
Asn Ala Phe Ser Leu Arg Val
    130                 135
```

<210> 243
<211> 110
<212> PRT
<213> Homo sapiens

<400> 243

```
Met Gly Ala Ala Met Ala Gly Gln Glu Asp Pro Val Gln Arg Glu Ile
 1                5                  10                  15
His Gln Asp Trp Ala Asn Arg Glu Tyr Ile Glu Ile Ile Thr Ser Ser
            20                  25                  30
Ile Lys Lys Ile Ala Asp Phe Leu Asn Ser Phe Asp Met Ser Cys Arg
        35                  40                  45
Ser Arg Leu Ala Thr Leu Asn Glu Lys Leu Thr Ala Leu Glu Arg Arg
    50                  55                  60
Ile Glu Tyr Ile Glu Ala Arg Val Thr Lys Gly Glu Thr Leu Thr Arg
65                  70                  75                  80
Thr Val Pro Cys Cys Cys Trp Glu Val Ala Leu His Asn Thr Gly His
                85                  90                  95
Met Gly Lys Ala Pro Ala Ala Phe Ser Ser Phe Leu Ser Pro
            100                 105                 110
```

<210> 244
<211> 122
<212> PRT
<213> Homo sapiens

<400> 244

```
Met Ala Ala Val Leu Gln Gln Val Leu Glu Asn Ala His Ile Lys Leu
 1                5                  10                  15
Ser Asn Leu Tyr Lys Ser Ala Ala Asp Asp Ser Glu Ala Lys Ser Asn
            20                  25                  30
Glu Leu Thr Arg Ala Val Glu Glu Leu His Lys Leu Leu Lys Glu Ala
        35                  40                  45
Gly Glu Ala Asn Lys Ala Ile Gln Asp His Leu Leu Glu Val Glu Gln
    50                  55                  60
Ser Lys Asp Gln Met Glu Lys Glu Met Leu Glu Lys Ile Gly Arg Leu
65                  70                  75                  80
Glu Lys Glu Leu Glu Asn Ala Asn Asp Leu Leu Ser Ala Thr Lys Arg
                85                  90                  95
Lys Gly Ala Ile Leu Ser Glu Glu Glu Leu Ala Ala Met Ser Pro Thr
            100                 105                 110
Arg Gly Gly Ile Asn Arg Gly Asn Ile Asn
        115                 120
```

<210> 245
<211> 19
<212> PRT
<213> Homo sapiens

<400> 245
Arg Trp Trp Met Cys Gly Gly Arg Met Glu Asp Met Leu Cys Cys Arg
1 5 10 15
Val Gly His

<210> 246
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 246
aactcctc 8

<210> 247
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 247
aatctacc 8

<210> 248
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 248
aacaacac 8

<210> 249
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 249
aatattcc 8

<210> 250
<211> 8

<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 250
ctcctaac 8

<210> 251
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 251
ctttctct 8

<210> 252
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 252
cttacttc 8

<210> 253
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Source Tag

<400> 253
atttcaat 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/06951

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G01N 33/53; C12P 21/06; C12N 1/20, 15/00, 15/09, 15/63, 15/70, 15/74; C07K 1/00, 14/00, 17/00; C07H 21/04

US CL : 435/7.1, 69.1, 252.3, 320.1, 325, 471; 530/350; 536/23.5

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/7.1, 69.1, 252.3, 320.1, 325, 471; 530/350; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,834,234 A (GALLO) 10 November 1998, Sequences 9 and 11 (attached Sequence Comparisons A and B).	1-8 and 14-19

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 October 2002 (24.10.2002)	11 DEC 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 ~~Facsimile No. (703)305-3230~~	Authorized officer Robert Landsman Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/06951

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-8 and 14-19 as drawn to SEQ ID NO:4 and 156

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/06951

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Gruops 1-75, claims 1-8, 14-19 in part, drawn to a human Bcl-XL-binding polypeptide, nucleic acid, vector, cell, and a method of identifying compounds that modulate Bcl-XL-binding polypeptide wherein the polypeptide is on of SEQ ID NO:4-71 and 224-230 and the nucleic acid is one of SEQ ID NO:156-223 and 231-237.

Group 76, claims 9-13, drawn to a method of identifying a BcL-XL-binding polypeptide.

Group 77, claim 20, drawn to a method of source-labeling a nucleic acid-protein fusion molecule.

Group 78, claim 21, drawn to a source-labeled nucleic acid -protein fusion molecule.

Group 79, claims 22-26, drawn to a method of identifying the source of a nucleic acid portion of a nucleic acid-protein fusion molecule.

The inventions listed as Groups 1-79 do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical feature for the following reasons: the special technical feature of Groups 1-75 are the 75 independent and distinct polypeptides of SEQ ID NO:4-71 and 224-230 and their corresponding nucleic acid molecules of SEQ ID NO:156-223 and 231-237. The special technical feature of Group 76 is a method of identifying a Bcl-XL-binding polypeptide. The special technical feature of Group 77 is a method of source-labeling a nucleic acid-protein fusion molecule. The special technical feature of Group 78 is a source-labeled nucleic acid-protein fusion molecule. The special technical feature of Group 79 is a method of identifying the source of the nucleic acid portion of a nucleic acid-protein fusion molecule. The special technical feature of each Group is not the same, or does not correspond to the special technical feature of any other Group because the products of Groups 1-75 and 78 are independent and distinct and the methods of Groups 76, 77 and 79 require different starting reagents and have different process steps and goals.

Continuation of B. FIELDS SEARCHED Item 3:

GenEmbl, N_Geneseq, A-Geneseq, Issued_Patents_AA, Issued_Patents_NA, PIR_71, SwissProt_40, SPTREMBL_19, EST